```
DEFINITION  Sequence 38 from Patent WO0078968.
ACCESSION   AX067463
VERSION     AX067463.1  GI:12545083
KEYWORDS    .
SOURCE      Moraxella catarrhalis.
  ORGANISM  Moraxella catarrhalis
            Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
            Moraxella.
REFERENCE   1  (bases 1 to 94750)
  AUTHORS   Lagace,R.E., Patterson,C. and Berg,K.L.
  TITLE     Nucleotide sequences of moraxella catarrhalis genome
  JOURNAL   Patent: WO 0078968-A 38 28-DEC-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .94750
                     /organism="Moraxella catarrhalis"
                     /db_xref="taxon:480"
BASE COUNT    27128 a  18811 c  21444 g  27367 t
ORIGIN


alignment_scores:
            Quality:  558.00              Length:    111
              Ratio:  5.027                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AX067463    ..

 Align seg 1/1  to: AX067463  from: 1  to: 94750

       1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75267 ATGTTAAATCATCATATTCGCCTGACTATTTCCGCTTTACTGACGGCACT 75316

      17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75317 TTTGGTAACAGGTTGTGTTTCTACTGGTAATGTTGCAATGAAAGAGCAAA 75366

      34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75367 ACCAACAAACCATTGAGCAGACCATCATTAAGGGCAAGACCAATAAGCAA 75416

      51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVa 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75417 GAGATTTCTAGTAGATTTGGTTCTGCTGATAGCATCTCTTTTATGATAGT 75466

      67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75467 GGTAATAAAATTTGGACATACCGCCATACTCGCTCCAAACCGATGGCAAG 75516

      84 luIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75517 AAATTTTATCCCTTATAATTTCTTTTCTTTGGGTGAAACCATACAGACCA 75566

     101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
         |||||||||||||||||||||||||||||||||
   75567 AAGAACTTGTCATTTTATTTGACAGCAAAGGCA 75599

seq_name: gb_ba:AE005493

seq_documentation_block:
LOCUS       AE005493               10701 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 112
            of 290.
ACCESSION   AE005493 AE005174
VERSION     AE005493.1  GI:12517067
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 10701)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 10701)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .10701
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .>10701
                     /note="O-island #108; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655"
     gene            complement(224. .829)
                     /gene="Z3935"
     CDS             complement(224. .829)
                     /gene="Z3935"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 82 to 187 of 201 are 30.84 pct identical
                     to residues 13 to 118 of 121 from GenPept 118 :
                     gi|15583|emb|CAA24404.1| (V01146) gene 3.8 [Bacteriophage
                     T7]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933Y"
                     /protein_id="AAG57748.1"
                     /db_xref="GI:12517068"
                     /translation="MAKFRKPLRRKKAMAKPARERCNRKREDLXVKRIFELLSFDKST
                     GVFRWKVPTQGRIALNSVAGAFDSNGYSMIMIDGRRYKTHVLVFYITHNRWPAGQIDH
                     VNGIRTDNRPENLRECLPIENSRNIRIRKNSKSGCRGVTWHKRQKKWNVRLGFHGKSK
                     HFGCFDDLELAVLVAEEARDKYYGDFSGNERSTYAMLSKEM"
     gene            1126. .1881
                     /gene="tra8_3"
                     /note="Z3936"
     CDS             1126. .1881
                     /gene="tra8_3"
                     /function="putative enzyme; Transposases (Insertion
                     Sequence Associated)"
                     /note="Residues 1 to 247 of 251 are 100.00 pct identical
                     to residues 1 to 247 of 383 from GenPept 118 :
                     gi|7800288|gb|AAF69884.1|AF250878_45 (AF250878) IS30
                     transposase [Salmonella typhi]"
                     /codon_start=1
                     /transl_table=11
                     /product="IS30 transposase"
                     /protein_id="AAG57749.1"
                     /db_xref="GI:12517069"
                     /translation="MRRTFTAEEKASVFELWKNGTGFSEIANILGSKPGTIFTMLRDT
                     GGIKPHERKRAVAHLTLSEREEIRAGLSAKMSIRAIATALNRSPSTISREVQRNRGRR
                     YYKAVDANNRANRMAKRPKPCLLDQNLPLRKLVLEKLEMKWSPEQISGWLRRTKPRQK
                     TLRISPETIYKTLYFRSREALHHLNIQHLRRSHSLRHGRRHTRKGERGTINIVNGTPI
                     HERSRNIDNRRSLGHWEGDLVSGTKNSHIDSFF"
     gene            1941. .2219
                     /gene="Z3937"
     CDS             1941. .2219
```

AA - MA

OM of: US-10-048-197-2 to: GenEmbl:* out_format : pfs

Date: Sep 18, 2002 5:46 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10048197/runat_17092002_135849_22601/app_query.fasta_1.168
-DB=GenEmbl -QFMT=fastap -SUFFIX=std.rge -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10048197_@CGN1_1_6377 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-10-048-197-2
Query length: 111
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 1887.650000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_pat:AX081233	+	558.00	1229.15	4.2e-60	336	AX081233 Sequence 1 from Patent
gb_pat:AX067463	+	558.00	1174.92	4.4e-57	94750	AX067463 Sequence 38 from Pat
gb_ba:AE005493	+	121.50	228.56	0.0002	10701	AE005493 Escherichia coli O15
gb_ba:AP002562	+	121.50	197.52	0.0122	270365	AP002562 Escherichia coli O1
gb_htg:AC101573	+	81.50	121.33	214.55	73989	AC101573 Mus musculus clone R
gb_htg:AC094388	+	81.50	115.13	475.06	140984	AC094388 Rattus norvegicus c
gb_htg:AL627314	+	81.50	111.93	715.97	196632	AL627314 Mus musculus chromo
gb_htg:AL627253	-	81.50	109.25	1.0e+03	259837	AL627253 Mus musculus chromo
gb_ba:AF169465	+	80.50	173.18	0.2775	267	AF169465 Neisseria meningitidis
gb_pat:A68918	+	80.50	173.18	0.2775	267	A68918 Sequence 89 from Patent
gb_pat:AX024066	+	80.00	169.78	0.4294	339	AX024066 Sequence 9 from Patent
gb_pat:AX033695	+	80.00	169.78	0.4294	339	AX033695 Sequence 1 from Patent
gb_in:CEB0491	-	79.00	122.13	193.50	38240	Z49907 Caenorhabditis elegans
gb_ba:NMA2Z2491	+	79.00	100.88	3.0e+03	349061	AL162753 Neisseria meningiti
gb_htg:AC102510	+	78.50	115.88	431.56	65312	AC102510 Mus musculus clone R
gb_pat:BD003755	+	78.00	134.80	38.15	8136	BD003755 Polynucleotide of Str
gb_ba:AE008532	-	78.00	132.44	51.59	10393	AE008532 Streptococcus pneumo
gb_ba:AE007477	-	78.00	130.41	66.96	12841	AE007477 Streptococcus pneumo
gb_pr:AC012068	-	78.00	105.78	1.6e+03	166516	AC012068 Homo sapiens BAC cl
gb_htg:AL356370	+	76.50	108.59	1.1e+03	87923	AL356370 Homo sapiens chromos
gb_htg:AL138848	+	76.50	107.13	1.3e+03	102351	AL138848 Homo sapiens chromo
gb_pr:AL162430	+	76.50	100.14	3.2e+03	211791	AL162430 Human DNA sequence
gb_pr:AL137850	+	75.50	102.93	2.3e+03	125842	AL137850 Human DNA sequence
gb_htg:AC011161	-	75.50	98.45	4.0e+03	200532	AC011161 Homo sapiens clone
gb_ov:AF173857	+	75.00	132.98	48.13	4920	AF173857 Xenopus laevis recept
gb_pr:AL512598	+	74.50	99.72	3.4e+03	139491	AL512598 Human DNA sequence
gb_htg:AP001795	-	74.50	99.42	3.6e+03	143873	AP001795 Homo sapiens chromo
gb_pl:AP003215	+	74.50	98.76	3.9e+03	154137	AP003215 Oryza sativa genomi
gb_htg:AC099643	-	74.50	97.73	4.4e+03	171660	AC099643 Mus musculus clone
gb_htg:AC021986	-	74.50	97.70	4.4e+03	172182	AC021986 Homo sapiens clone
gb_pr:AC019184	+	74.50	96.53	5.2e+03	194372	AC019184 Homo sapiens BAC cl
gb_pr:AC062004	-	74.50	95.99	5.5e+03	205646	AC062004 Homo sapiens chromo
gb_ba:AF253043	+	74.00	143.01	13.30	1377	AF253043 Lactobacillus crispat
gb_ba:AF421351	+	74.00	135.90	33.10	2884	AF421351 Azotobacter vinelandi
gb_htg:AC067965	+	74.00	106.08	1.5e+03	64165	AC067965 Arabidopsis thaliana
gb_htg:SPNEU1902	-	74.00	97.79	4.4e+03	151947	AL449924 Streptococcus pneum
gb_htg:AC009452	+	74.00	96.46	5.2e+03	174431	AC009452 Homo sapiens chromo
gb_htg:AC099358	-	74.00	96.30	5.3e+03	177495	AC099358 Rattus norvegicus c
gb_pr:AL137020	+	74.00	96.04	5.5e+03	182231	AL137020 Human DNA sequence
gb_htg:AC099386	-	74.00	94.36	6.8e+03	217121	AC099386 Rattus norvegicus c
gb_htg:AC106605	-	74.00	92.22	9.0e+03	271135	AC106605 Rattus norvegicu
gb_ba:AP003589	+	74.00	90.00	1.2e+04	341880	AP003589 Nostoc sp. PCC 7
gb_pl:ATT4D2	+	73.50	101.44	2.7e+03	92611	AL132958 Arabidopsis thali
gb_htg:AC106635	+	73.50	97.78	4.4e+03	135538	AC106635 Rattus norvegicu
gb_htg:AP003572	+	73.50	95.62	5.8e+03	169761	AP003572 Oryza sativa chr

seq_name: gb_pat:AX081233

seq_documentation_block:
LOCUS AX081233 336 bp DNA linear PAT 27-FEB-2001
DEFINITION Sequence 1 from Patent WO0109337.
ACCESSION AX081233
VERSION AX081233.1 GI:13170101
KEYWORDS .
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 336)
AUTHORS Thonnard,J.S.
TITLE Moraxella catarrhalis antigens basb122 and basb124
JOURNAL Patent: WO 0109337-A 1 08-FEB-2001;
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES Location/Qualifiers
source 1. .336
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 106 a 65 c 61 g 104 t
ORIGIN

alignment_scores:
Quality: 558.00 Length: 111
Ratio: 5.027 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AX081233 ..

Align seg 1/1 to: AX081233 from: 1 to: 336

```
      1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
      1 ATGTTAAATCATCATATTCGCCTGACTATTTCCGCTTTACTGACGGCACT 50

     17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnA 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
     51 TTTGGTAACAGGTTGTGTTTCTACTGGTAATGTTGCAATGAAAGAGCAAA 100

     34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    101 ACCAACAAACCATTGAGCAGACCATCATTAAGGGCAAGACCAATAAGCAA 150

     51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVa 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    151 GAGATTTCTAGTAGATTTGGTTCTGCTGATAGCATCTCTTTTATGATAGT 200

     67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnG 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    201 GGTAATAAAATTTGGACATACCGCCATACTCGCTCCAAACCGATGGCAAG 250

     84 luIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
    251 AAATTTTATCCCTTATAATTTCTTTTCTTTGGGTGAAACCATACAGACCA 300

    101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
        |||||||||||||||||||||||||||||||||
    301 AAGAACTTGTCATTTTATTTGACAGCAAAGGCA 333
```

seq_name: gb_pat:AX067463

seq_documentation_block:
LOCUS AX067463 94750 bp DNA linear PAT 24-JAN-2001

```
DEFINITION  Sequence 38 from Patent WO0078968.
ACCESSION   AX067463
VERSION     AX067463.1  GI:12545083
KEYWORDS    .
SOURCE      Moraxella catarrhalis.
  ORGANISM  Moraxella catarrhalis
            Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
            Moraxella.
REFERENCE   1  (bases 1 to 94750)
  AUTHORS   Lagace,R.E., Patterson,C. and Berg,K.L.
  TITLE     Nucleotide sequences of moraxella catarrhalis genome
  JOURNAL   Patent: WO 0078968-A 38 28-DEC-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .94750
                     /organism="Moraxella catarrhalis"
                     /db_xref="taxon:480"
BASE COUNT    27128 a  18811 c  21444 g  27367 t
ORIGIN


alignment_scores:
            Quality:  558.00              Length:    111
              Ratio:  5.027                Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AX067463   ..

Align seg 1/1  to: AX067463  from: 1  to: 94750

      1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  75267 ATGTTAAATCATCATATTCGCCTGACTATTTCCGCTTTACTGACGGCACT 75316

     17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnA 34
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  75317 TTTGGTAACAGGTTGTGTTTCTACTGGTAATGTTGCAATGAAAGAGCAAA 75366

     34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  75367 ACCAACAAACCATTGAGCAGACCATCATTAAGGGCAAGACCAATAAGCAA 75416

     51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVa 67
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  75417 GAGATTTCTAGTAGATTTGGTTCTGCTGATAGCATCTCTTTTATGATAGT 75466

     67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnG 84
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  75467 GGTAATAAAATTTGGACATACCGCCATACTCGCTCCAAACCGATGGCAAG 75516

     84 luIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
        ||||||||||||||||||||||||||||||||||||||||||||||||||
  75517 AAATTTTATCCCTTATAATTTCTTTTCTTTGGGTGAAACCATACAGACCA 75566

    101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
        |||||||||||||||||||||||||||||||
  75567 AAGAACTTGTCATTTTATTTGACAGCAAAGGCA 75599

seq_name: gb_ba:AE005493

seq_documentation_block:
LOCUS       AE005493               10701 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 112
            of 290.
ACCESSION   AE005493 AE005174
VERSION     AE005493.1  GI:12517067
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 10701)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 10701)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .10701
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     misc_feature    <1. .>10701
                     /note="O-island #108; Region of the EDL933 chromosome not
                     homologous to E. coli K-12 MG1655"
     gene            complement(224. .829)
                     /gene="Z3935"
     CDS             complement(224. .829)
                     /gene="Z3935"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="Residues 82 to 187 of 201 are 30.84 pct identical
                     to residues 13 to 118 of 121 from GenPept 118 :
                     gi|15583|emb|CAA24404.1| (V01146) gene 3.8 [Bacteriophage
                     T7]"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown protein encoded by prophage CP-933Y"
                     /protein_id="AAG57748.1"
                     /db_xref="GI:12517068"
                     /translation="MAKFRKPLRRKKAMAKPARRRCNRKREDLXVKRIFELLSFDKST
                     GVFRWKVPTQGRIALNSVAGAFDSNGYSMIMIDGRRYKTHVLVFYITHNRWPAGQIDH
                     VNGIRTDNRPENLRECLPIENSRNIRIRKNSKSGCRGVTWHKRQKKWNVRLGFHGKSK
                     HFGCFDDLELAVLVAEEARDKYYGDFSGNERSTYANLSKEM"
     gene            1126. .1881
                     /gene="tra8_3"
                     /note="Z3936"
     CDS             1126. .1881
                     /gene="tra8_3"
                     /function="putative enzyme; Transposases (Insertion
                     Sequence Associated)"
                     /note="Residues 1 to 247 of 251 are 100.00 pct identical
                     to residues 1 to 247 of 383 from GenPept 118 :
                     gi|7800288|gb|AAF69884.1|AF250878_45 (AF250878) IS30
                     transposase [Salmonella typhi]"
                     /codon_start=1
                     /transl_table=11
                     /product="IS30 transposase"
                     /protein_id="AAG57749.1"
                     /db_xref="GI:12517069"
                     /translation="MRRTFTAEEKASVFELWKNGTGFSEIANILGSKPGTIFTMLRDT
                     GGIKPHERKRAVAHLTLSEREEIRAGLSAKMSIRAIATALNRSPSTISREVQRNRGRR
                     YYKAVDANNRANRMAKRPKPCLLDQNLPLRKLVLEKLEMKWSPEQISGWLRRTKPRQK
                     TLRISPETIYKTLYFRSREALHHLNIQHLRRSHSLRHGRRHTRKGERGTINIVNGTPI
                     HERSRNIDNRRSLGHWEGDLVSGTKNSHIDSFF"
     gene            1941. .2219
                     /gene="Z3937"
     CDS             1941. .2219
```

```
                     /gene="Z3937"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAG57750.1"
                     /db_xref="GI:12517070"
                     /translation="MKTKIVKGKTTKQDVLASFGEPDSRSLIDGEEQWSYTMYNSQSK
                     ATSFIPVVGLLAGGADSQTKSLTVSFKGEKVSTYIFNAGTSNVKTGIF"
     gene            2390. .2509
                     /gene="Z3938"
     CDS             2390. .2509
                     /gene="Z3938"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAG57751.1"
                     /db_xref="GI:12517071"
                     /translation="MALALALALALALALALALALALALALALALALALAKEN"
     gene            complement(2609. .3517)
                     /gene="Z3939"
     CDS             complement(2609. .3517)
                     /gene="Z3939"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAG57752.1"
                     /db_xref="GI:12517072"
                     /translation="MXIAKDSSTDEIIHGNDLRGMDDFYIKTTSFECPYEPCKIKATP
                     CSFTMKHVNQSYFRYGDKHKDGCGIHDPRYKNNHTSNDERKHNSPPAPVISLLKIDVK
                     PRGGVKNARNIKNENHKDEKKGNEHPVSSSSIKPVVDYYINNSNHNEQLSIPPYGTRS
                     YKDTFQLIFYKNNIRYYKPAIYYGVAQSNIRLHEDSDKHCITFLARDKKTQKPFTLEI
                     DVSDWNKSQKDVFWKEYEKQRKEADRYYKGLKDKRNAKKYLTVFFFGMPDENDKFLFK
                     TNHFKLVYVAFLGKFESSYNDSNYYH"
     gene            complement(3914. .4384)
                     /gene="Z3940"
     CDS             complement(3914. .4384)
                     /gene="Z3940"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAG57753.1"
                     /db_xref="GI:12517073"
                     /translation="MLFIFLSGRQNFTIARMHIMDKTNINLMERYLLLLDRFVDKLAE
                     SGFSEQRIIEQSYLFCAGFYIKYQSGIEKMTFSNREVVLTFLLLSYYSHINKLDDDLI
                     NKERMKHVCSSLINFIVSNGSRTEKVYANEKRKYEASTLKRELSKKDKRKRYRL"
     gene            complement(4425. .4646)
                     /gene="Z3941"
     CDS             complement(4425. .4646)
                     /gene="Z3941"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAG57754.1"
                     /db_xref="GI:12517074"
                     /translation="MKKTTIKSGVIHTVKPANLKVASQRIKRLEAENEALKALNERYL
                     EQFKVWLYNAHLKQITIEELNNPLPAKYL"
     gene            complement(5046. .6458)
                     /gene="Z3942"
     CDS             complement(5046. .6458)
                     /gene="Z3942"
                     /function="orf; Other or unknown (Phage or Prophage
                     Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAG57755.1"
                     /db_xref="GI:12517075"
                     /translation="MNNIFIFEPSNKNNPLDNVIKFIEFCKSTISNNNLTTSWESNKW
                     KGLYRFTKFNSKNNLNSKECLDDSFINFAKAYMLHVHSFNKSKTKHSTLSMLKIVEFV
                     LLKINMEANVNYCNNSIYDECIRIASEKYSKAHAFAIGKELEKLSSFLNDNRMTNSFY
                     LFWVNPIRYRITQSWTGYDSSLEGHSRLPDIKSVIAIAEIFSKRDEQLSSRDIFTTSV
                     LALLMCAPSRISEILALPADCEITECDGKGIQRYGLRFFSAKGYEGNIKWIPTLMIPV
                     AKKAISRLKELSSQARLLAAEIQKNYSNSTKGTLKENIPPDLFWYDREKKIKYSNALC
                     LLTEGQLNQNKKEMSDKLFRPTTNFFKTDIIDSDYIKGYFNVFKRHGYINEDGSPYLL
                     XTHQLRHLLNTFAXINGMDEFSIARWSGRKLISQNVSYDHRSHLQMSKAIREKKLSVC
                     VNEHRIKDIPVVDLNEFGLT"
     gene            complement(6455. .7780)
                     /gene="Z3943"
     CDS             complement(6455. .7780)
                     /gene="Z3943"
                     /function="putative enzyme; Integration, recombination
                     (Phage or Prophage Related)"
                     /note="No significant matches"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAG57756.1"
                     /db_xref="GI:12517076"
                     /translation="MEFSMENKCIESEQIFFAKMNRYSFKLSDKKWQLDKENCVYPHK
                     VVDRMPTKMKLSYLKTLAYYASEYSSFYIQSINNLFYEWFGAMTIDTIDDKAIYQLNV
                     YLGSERNYKLNLIKAFIIKWKNLNYPGVEATAIRMLEKIKIIPNQTGDAVKRRDPNKG
                     PLTEAEFNNIINAVGKFYHEKKIQCFLYCYILLLAITGRRPLQLISLKAKDLIKNERG
                     CFLNVPKVKQRKCFRKEFNMVMIEPFLYDSLSMLINQNQAFVEDKFSVGISNYRGELP
                     IFMNLDKITETKRIEDFLYDLTTDFFHMKNSVMSKLLKHFPSKFDVRSERTNSYIELN
                     ARRFRYTLGSRLANEGASIEVIAKALDHKSVNSSIIYIKNNPDNVYDIDKRLSAFFNP
                     LSNILMGIEIEENKNFFIKFVSDAFFLLEDTKEDLKCLTCKKFNPWRAL"
     gene            complement(7935. .8978)
                     /gene="Z3945"
     CDS             complement(7935. .8978)
                     /gene="Z3945"
                     /function="putative enzyme; Integration, recombination
                     (Phage or Prophage Related)"
                     /note="Residues 172 to 336 of 347 are 28.57 pct identical
                     to residues 125 to 264 of 311 from GenPept 118 :
                     gi|2621988|gb|AAB85391.1| (AE000865) integrase-recombinase
                     protein [Methanobacterium thermoautotrophicum]"
                     /codon_start=1

alignment_scores:
           Quality:  121.50            Length:     93
             Ratio:  1.960               Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 34.409

alignment_block:
 US-10-048-197-2 x AE005493   ..

 Align seg 1/1  to: AE005493  from: 1  to: 10701

      17 LeuLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGl 33
         |||   ::::::||||||   |||:::||||||   :::::::|||:::::
    1878 CTGACCTTGGCAGGTTGTTCTTCAGTAGGCAACCAGTCATTGAAAAATGA 1927

      33 nAsnGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysG 50
         ::::|||:::::::::::::      |||:::|||||||||||||:::||||
    1928 GACGCAGGAAAGTGTGAAAACCAAAATTGTTAAAGGCAAAACTACAAAAC 1977

      50 lnGluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIle 66
         ||::::::   :::   ||||||||:::   ||||||   |||   :::|||
    1978 AGGACGTGTTAGCATCGTTCGGTGAACCTGACAGCCGTTCT...TTGATC 2024

      67 .ValValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpG 83
          :::|||   :::   ||||||||      ::::::   :::       |
    2025 GATGGTGAAGAACAATGGTCATACACTATGTATAACAGCCAGTCCAAAGC 2074

      83 lnGluIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArg 99
         ||   :::   |||   :::   :::   |||   |||         :::
    2075 AACCTCTTTCATCCCCGTTGTGGGACTGCTTGCAGGTGGCGCAGACTCAC 2124
```

```
      100 ProLysAsnLeuSerPheTyrLeuThr 108
                ||||||      :::|||:::
     2125 AAACTAAATCTCTGACAGTTTCTTTCA 2151

seq_name: gb_ba:AP002562

seq_documentation_block:
LOCUS       AP002562               270365 bp    DNA     linear   BCT 07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 13/20.
ACCESSION   AP002562 BA000007
VERSION     AP002562.1  GI:13362858
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 270365)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .270365
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            complement(435. .1220)
                     /gene="ECs3388"
     CDS             complement(435. .1220)
                     /gene="ECs3388"
                     /note="similar to SSEB_ECOLI gi|1788871 percent identity
                     98 in 261 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="enhanced serine sensitivity"
                     /protein_id="BAB36811.1"
                     /db_xref="GI:13362859"
                     /translation="MVLMSETKNELEDLLEKAATEPAHRPAFFRTLLESTVWVPGTAA
                     QGEAVVEDSALDLQHWEKEDGTSVIPFFTSLEALQQAVEDEQAFVVMPVRTLFEMTLG
                     ETLFLNAKLPTGKEFMPREISLLIGEEGNPLSSQEILEGGESLILSEVAEPPAQMIDS
                     LTTLFKTIKPVKRAFICSIKENEEAQPNLLIGIEADGDIEEIIQVAGSVATDTLPGDE
                     PIDICQVKKGEKGISHFITEHIAPFYERRWGGFLRDFKQNRII"
     gene            complement(1354. .2724)
                     /gene="ECs3389"
     CDS             complement(1354. .2724)
                     /gene="ECs3389"
                     /note="similar to PEPB_ECOLI gi|1788872 percent identity
                     99 in 456 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative peptidase"
                     /protein_id="BAB36812.1"
                     /db_xref="GI:13362860"
                     /translation="MPPSCNANHLAEKSWGCQLAALLQRRITKMTEAMKITLSTQPAD
                     ARWGEKATYSINNDGITLHLNGADDLGLIQRAARKIDGLGIKHVQLSGEGWDADRCWA
                     FWQGYKAPKGTRKVEWPDLDDAQRQELDNRLMIIDWVRDTINAPAEELGPSQLAQRAV
                     DLISNVASDRVTYRITKGEDLREQGYMGLHTVGRGSERSPVLLALDYNPTGDKEAPVY
                     ACLVGKGITFDSGGYSIKQTAFMDSMKSDMGGAATVTGALAFAITRGLNKRVKLFLCC
                     ADNLISGNAFKLGDIITYRNGKKVEVMNTDAEGRLVLADGLIDASAQKPEMIIDAATL
                     TGAAKTALGNDYHALFSFDDALAGRLLASASQENEPFWRLPLAEFHRSQLPSNFAELN
                     NTGSAAYPAGASTAAGFLSHFVENYQQGWLHIDCSATYRKAPVEQWSAGATGLGVRTI
                     ANLLTA"
     gene            complement(2815. .3015)
                     /gene="ECs3390"
     CDS             complement(2815. .3015)
                     /gene="ECs3390"
                     /note="identical to YFHJ_ECOLI gi|1788873 (Conserved in
                     E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="hypothetical protein"
                     /protein_id="BAB36813.1"
                     /db_xref="GI:13362861"
                     /translation="MGLKWTDSREIGEALYDAYPDLDPKTVRFTDMHQWICDLEDFDD
                     DPQASNEKILEAILLVWLDEAE"
     gene            complement(3027. .3362)
                     /gene="ECs3391"
     CDS             complement(3027. .3362)
                     /gene="ECs3391"
                     /note="similar to FER_ECOLI gi|119999|sp|P25528 percent
                     identity 100 in 111 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="[2FE-2S] ferredoxin"
                     /protein_id="BAB36814.1"
                     /db_xref="GI:13362862"
                     /translation="MPKIVILPHQDLCPDGAVLEANSGETILDAALRNGIEIEHACEK
                     SCACTTCHCIVREGFDSLPESSEQEDDMLDKAWGLEPESRLSCQARVTDEDLVVEIPR
                     YTINHAREH"
     gene            complement(3364. .5214)
                     /gene="ECs3392"
     CDS             complement(3364. .5214)
                     /gene="ECs3392"
                     /note="similar to HSCA_ECOLI gi|1788875 percent identity
                     100 in 616 aa (Conserved in E.coli K-12)"
```

```
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="molecular chaperone"
                     /protein_id="BAB36815.1"
                     /db_xref="GI:13362863"
                     /translation="MALLQISEPGLSAAPHQRRLAAGIDLGTTNSLVATVRSGQAETL
                     ADHEGRHLLPSVVHYQQQGHSVGYDARTNAALDTANTISSVKRLMGRSLADIQQRYPH
                     LPYQFQASENGLPMIETAAGLLNPVRVSADILKALAARATEALAGELDGVVITVPAYF
                     DDAQRQGTKDAARLAGLHVLRLLNEPTAAAIAYGLDSGQEGVIAVYDLGGGTFDISIL
                     RLSRGVFEVLATGGDSALGGDDFDHLLADYIREQAGIPDRSDNRVQRELLDAAIAAKI
                     ALSDADSVTVNVAGWQGEISREQFNELIAPLVKRTLLACRRALKDAGVEADEVLEVVM
                     VGGSTRVPLVRERVGEFFGRPPLTSIDPDKVVAIGAAIQADILVGNKPDSEMLLLDVI
                     PLSLGLETMGGLVEKVIPRNTTIPVARAQDFTTFKDGQTAMSIHVMQGERELVQDCRS
                     LARFALRGIPALPAGGAHIRVTFQVDADGLLSVTAMEKSTGVEASIQVKPSYGLTDSE
                     IASMIKDSMSYAEQDVKARMLAEQKVEAARVLESLHGALAADAALLSAAERQVIDDAA
                     AHLSEVAQGDDVDAIEQAIKNVDKQTQDFAARRMDQSVRRALKGHSVDEV"
     gene            complement(5231. .5746)
                     /gene="ECs3393"
     CDS             complement(5231. .5746)
                     /gene="ECs3393"
                     /note="similar to HSCB_ECOLI gi|1788876 percent identity
                     100 in 171 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="molecular chaperone"
                     /protein_id="BAB36816.1"
                     /db_xref="GI:13362864"
                     /translation="MDYFTLFGLPARYQLDTQALSLRFQDLQRQYHPDKFASGSQAEQ
                     LAAVQQSATINQAWQTLRHPLMRAEYLLSLHGFDLASEQHTVRDTAFLMEQLELREEL
                     DEIEQAKDEARLESFIKRVKKMFDTRHQLMVEQLDNETWDAAADTVRKLRFLDKLRSS
                     AEQLEEKLLDF"
     gene            complement(5842. .6165)
                     /gene="ECs3394"
     CDS             complement(5842. .6165)
                     /gene="ECs3394"
                     /note="similar to ISCA_ECOLI gi|1788877 percent identity
                     100 in 107 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="putative iron-binding protein"
                     /protein_id="BAB36817.1"
                     /db_xref="GI:13362865"
                     /translation="MSITLSDSAAARVNTFLANRGKGFGLRLGVRTSGCSGMAYVLEF
                     VDEPTPEDIVFEDKGVKVVVDGKSLQFLDGTQLDFVKEGLNEGFKFTNPNVKDECGCG
                     ESFHV"
     gene            complement(6182. .6568)
                     /gene="ECs3395"
     CDS             complement(6182. .6568)
                     /gene="ECs3395"
                     /note="similar to ISCU_ECOLI gi|7387937|sp|P77310 percent
                     identity 100 in 128 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /transl_table=11
                     /evidence=not_experimental
                     /product="NifU-like protein"
                     /protein_id="BAB36818.1"
                     /db_xref="GI:13362866"
                     /translation="MAYSEKVIDHYENPRNVGSFDNNDENVGSGMVGAPACGDVMKLQ
                     IKVNDEGIIEDARFKTYGCGSAIASSSLVTEWVKGKSLDEAQAIKNTDIAEELELPPV

alignment_scores:
           Quality:  121.50            Length:     93
             Ratio:  1.960               Gaps:      2
 Percent Similarity:  66.667   Percent Identity: 34.409

alignment_block:
 US-10-048-197-2 x AP002562   ..

 Align seg 1/1  to: AP002562  from: 1  to: 270365

      17 LeuLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGl 33
         |||   ::::::||||||   |||:::||||||   ::::::|||:::::
  116902 CTGACCTTGGCAGGTTGTTCTTCAGTAGGCAACCAGTCATTGAAAAATGA 116951

      33 nAsnGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysG 50
         :::::|||:::::::::::::       |||:::|||||||||||||:::||||
  116952 GACGCAGGAAAGTGTGAAAACCAAAATTGTTAAAGGCAAAACTACAAAAC 117001

      50 lnGluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIle 66
         ||:::::   :::   ||||||:::   ||||||   |||   :::|||
  117002 AGGACGTGTTAGCATCGTTCGGTGAACCTGACAGCCGTTCT...TTGATC 117048

      67 .ValValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpG 83
          :::|||   :::   |||||||||       ::::::   :::       |
  117049 GATGGTGAAGAACAATGGTCATACACTATGTATAACAGCCAGTCCAAAGC 117098

      83 lnGluIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArg 99
         ||   :::   |||   :::   :::   |||   |||         :::
  117099 AACCTCTTTCATCCCCGTTGTGGGACTGCTTGCAGGTGGCGCAGACTCAC 117148

     100 ProLysAsnLeuSerPheTyrLeuThr 108
               ||||||      :::|||:::
  117149 AAACTAAATCTCTGACAGTTTCTTTCA 117175

seq_name: gb_htg:AC101573

seq_documentation_block:
LOCUS       AC101573              73989 bp    DNA     linear   HTG 23-NOV-2001
DEFINITION  Mus musculus clone RP23-197A22, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC101573
VERSION     AC101573.1  GI:17060348
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 73989)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-197A22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 73989)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
```

Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L17015
Center clone name: 197_A_22
----------.
* NOTE: This record contains 92 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 719: contig of 719 bp in length
* 720 819: gap of 100 bp
* 820 1536: contig of 717 bp in length
* 1537 1636: gap of 100 bp
* 1637 2339: contig of 703 bp in length
* 2340 2439: gap of 100 bp
* 2440 3132: contig of 693 bp in length
* 3133 3232: gap of 100 bp
* 3233 3934: contig of 702 bp in length
* 3935 4034: gap of 100 bp
* 4035 4745: contig of 711 bp in length
* 4746 4845: gap of 100 bp
* 4846 5546: contig of 701 bp in length
* 5547 5646: gap of 100 bp
* 5647 6348: contig of 702 bp in length
* 6349 6448: gap of 100 bp
* 6449 7166: contig of 718 bp in length
* 7167 7266: gap of 100 bp
* 7267 7983: contig of 717 bp in length
* 7984 8083: gap of 100 bp
* 8084 8786: contig of 703 bp in length
* 8787 8886: gap of 100 bp
* 8887 9605: contig of 719 bp in length
* 9606 9705: gap of 100 bp
* 9706 10396: contig of 691 bp in length
* 10397 10496: gap of 100 bp
* 10497 11200: contig of 704 bp in length
* 11201 11300: gap of 100 bp
* 11301 12012: contig of 712 bp in length
* 12013 12112: gap of 100 bp
* 12113 12815: contig of 703 bp in length
* 12816 12915: gap of 100 bp
* 12916 13614: contig of 699 bp in length
* 13615 13714: gap of 100 bp
* 13715 14428: contig of 714 bp in length
* 14429 14528: gap of 100 bp
* 14529 15243: contig of 715 bp in length
* 15244 15343: gap of 100 bp
* 15344 16072: contig of 729 bp in length
* 16073 16172: gap of 100 bp
* 16173 16890: contig of 718 bp in length
* 16891 16990: gap of 100 bp
* 16991 17695: contig of 705 bp in length
* 17696 17795: gap of 100 bp
* 17796 18482: contig of 687 bp in length
* 18483 18582: gap of 100 bp
* 18583 19294: contig of 712 bp in length
* 19295 19394: gap of 100 bp
* 19395 20106: contig of 712 bp in length
* 20107 20206: gap of 100 bp
* 20207 20911: contig of 705 bp in length
* 20912 21011: gap of 100 bp
* 21012 21711: contig of 700 bp in length
* 21712 21811: gap of 100 bp
* 21812 22516: contig of 705 bp in length
* 22517 22616: gap of 100 bp
* 22617 23310: contig of 694 bp in length
* 23311 23410: gap of 100 bp
* 23411 24116: contig of 706 bp in length
* 24117 24216: gap of 100 bp
* 24217 24921: contig of 705 bp in length
* 24922 25021: gap of 100 bp
* 25022 25734: contig of 713 bp in length
* 25735 25834: gap of 100 bp
* 25835 26548: contig of 714 bp in length
* 26549 26648: gap of 100 bp
* 26649 27355: contig of 707 bp in length
* 27356 27455: gap of 100 bp
* 27456 28147: contig of 692 bp in length
* 28148 28247: gap of 100 bp
* 28248 28959: contig of 712 bp in length
* 28960 29059: gap of 100 bp
* 29060 29755: contig of 696 bp in length
* 29756 29855: gap of 100 bp
* 29856 30561: contig of 706 bp in length
* 30562 30661: gap of 100 bp
* 30662 31372: contig of 711 bp in length
* 31373 31472: gap of 100 bp
* 31473 32164: contig of 692 bp in length
* 32165 32264: gap of 100 bp
* 32265 32940: contig of 676 bp in length
* 32941 33040: gap of 100 bp
* 33041 33757: contig of 717 bp in length
* 33758 33857: gap of 100 bp
* 33858 34568: contig of 711 bp in length
* 34569 34668: gap of 100 bp
* 34669 35379: contig of 711 bp in length
* 35380 35479: gap of 100 bp
* 35480 36198: contig of 719 bp in length
* 36199 36298: gap of 100 bp
* 36299 37004: contig of 706 bp in length
* 37005 37104: gap of 100 bp
* 37105 37810: contig of 706 bp in length
* 37811 37910: gap of 100 bp
* 37911 38582: contig of 672 bp in length
* 38583 38682: gap of 100 bp
* 38683 39384: contig of 702 bp in length
* 39385 39484: gap of 100 bp
* 39485 40168: contig of 684 bp in length
* 40169 40268: gap of 100 bp
* 40269 40963: contig of 695 bp in length
* 40964 41063: gap of 100 bp
* 41064 41765: contig of 702 bp in length
* 41766 41865: gap of 100 bp
* 41866 42570: contig of 705 bp in length
* 42571 42670: gap of 100 bp
* 42671 43368: contig of 698 bp in length
* 43369 43468: gap of 100 bp
* 43469 44188: contig of 720 bp in length
* 44189 44288: gap of 100 bp
* 44289 44984: contig of 696 bp in length
* 44985 45084: gap of 100 bp
* 45085 45801: contig of 717 bp in length
* 45802 45901: gap of 100 bp
* 45902 46588: contig of 687 bp in length
* 46589 46688: gap of 100 bp
* 46689 47360: contig of 672 bp in length
* 47361 47460: gap of 100 bp
* 47461 48180: contig of 720 bp in length
* 48181 48280: gap of 100 bp
* 48281 48967: contig of 687 bp in length
* 48968 49067: gap of 100 bp
* 49068 49770: contig of 703 bp in length
* 49771 49870: gap of 100 bp
* 49871 50580: contig of 710 bp in length
* 50581 50680: gap of 100 bp
* 50681 51408: contig of 728 bp in length
* 51409 51508: gap of 100 bp
* 51509 52241: contig of 733 bp in length
* 52242 52341: gap of 100 bp

```
            *    52342    53050: contig of 709 bp in length
            *    53051 53150: gap of      100 bp
            *    53151    53865: contig of 715 bp in length
            *    53866 53965: gap of      100 bp
            *    53966    54654: contig of 689 bp in length
            *    54655 54754: gap of      100 bp
            *    54755    55466: contig of 712 bp in length
            *    55467 55566: gap of      100 bp

alignment_scores:
           Quality:   81.50              Length:     77
             Ratio:  1.598                Gaps:      1
 Percent Similarity:  66.234   Percent Identity: 27.273

alignment_block:
 US-10-048-197-2 x AC101573   ..

 Align seg 1/1  to: AC101573  from: 1  to: 73989

      31 LysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLysGlyLysTh 47
         |||:::::::||||||   ::::::          ::::::   :::   ::
   61496 AAAAAAAAAAACCAAACCTCCATGCCTTTCCCTTTCTTTCTCTCTTGCTC 61545

      47 rAsnLysGlnGluIleSerSerArgPheGlySerAlaAspSerIleSerP 64
         :|||:::   ::::::||||||   |||       :::   |||:::   |
   61546 TAATTCGCCCTCACTCTCTTCCTTCTTTCCTCCCTCTCCCTCTCTTTGTT 61595

      64 heMetIleValValIleLysPheGlyHisThrAlaIleLeuAlaProAsn 80
         ||   :::   ::::::::::|||       :::   :::|||:::|||:::
   61596 TCTGTTTGTTTCTTCTTTCCTTTCTCCTTTCTTTCTTCCTTTCTCCTTCT 61645

      81 ArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrpValLysPr 97
                :::::::||||||      :::||||||||||:::::   ||
   61646 .........TCCTTCCTTTCTTCCTCTCTTTCCTTCCTTTTCTTCCTCC 61686

      97 oTyrArgProLysAsnLeuSerPheTyrLeu 107
         |:::   |||:::::::|||      :::|||
   61687 CTTCCTCCCTTCCTCCCTTCCTCCCTTCCTC 61717

seq_name: gb_htg:AC094388

seq_documentation_block:
LOCUS       AC094388              140984 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-3N14, *** SEQUENCING IN PROGRESS ***,
            61 unordered pieces.
ACCESSION   AC094388
VERSION     AC094388.2  GI:17941111
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 140984)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 140984)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:15624222.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GAMS
                Center clone name: CH230-3N14
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 112041 bases at least Q40
                Consensus quality: 120432 bases at least Q30
                Consensus quality: 127286 bases at least Q20
                Estimated insert size: 121178; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 61 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3813: contig of 3813 bp in length
            *     3814     3913: gap of unknown length
            *     3914     8564: contig of 4651 bp in length
            *     8565     8664: gap of unknown length
            *     8665    12409: contig of 3745 bp in length
            *    12410    12509: gap of unknown length
            *    12510    16844: contig of 4335 bp in length
            *    16845    16944: gap of unknown length
            *    16945    21435: contig of 4491 bp in length
            *    21436    21535: gap of unknown length
            *    21536    23974: contig of 2439 bp in length
            *    23975    24074: gap of unknown length
            *    24075    28394: contig of 4320 bp in length
            *    28395    28494: gap of unknown length
            *    28495    33266: contig of 4772 bp in length
            *    33267    33366: gap of unknown length
```

```
*    33367    36620: contig of 3254 bp in length
*    36621    36720: gap of unknown length
*    36721    39615: contig of 2895 bp in length
*    39616    39715: gap of unknown length
*    39716    42700: contig of 2985 bp in length
*    42701    42800: gap of unknown length
*    42801    46577: contig of 3777 bp in length
*    46578    46677: gap of unknown length
*    46678    48925: contig of 2248 bp in length
*    48926    49025: gap of unknown length
*    49026    51536: contig of 2511 bp in length
*    51537    51636: gap of unknown length
*    51637    54619: contig of 2983 bp in length
*    54620    54719: gap of unknown length
*    54720    57075: contig of 2356 bp in length
*    57076    57175: gap of unknown length
*    57176    59447: contig of 2272 bp in length
*    59448    59547: gap of unknown length
*    59548    63145: contig of 3598 bp in length
*    63146    63245: gap of unknown length
*    63246    66267: contig of 3022 bp in length
*    66268    66367: gap of unknown length
*    66368    69215: contig of 2848 bp in length
*    69216    69315: gap of unknown length
*    69316    72092: contig of 2777 bp in length
*    72093    72192: gap of unknown length
*    72193    73295: contig of 1103 bp in length
*    73296    73395: gap of unknown length
*    73396    75943: contig of 2548 bp in length
*    75944    76043: gap of unknown length
*    76044    78936: contig of 2893 bp in length
*    78937    79036: gap of unknown length
*    79037    80901: contig of 1865 bp in length
*    80902    81001: gap of unknown length
*    81002    83188: contig of 2187 bp in length
*    83189    83288: gap of unknown length
*    83289    85976: contig of 2688 bp in length
*    85977    86076: gap of unknown length
*    86077    87770: contig of 1694 bp in length
*    87771    87870: gap of unknown length
*    87871    88938: contig of 1068 bp in length
*    88939    89038: gap of unknown length
*    89039    91036: contig of 1998 bp in length
*    91037    91136: gap of unknown length
*    91137    93364: contig of 2228 bp in length
*    93365    93464: gap of unknown length
*    93465    94847: contig of 1383 bp in length
*    94848    94947: gap of unknown length
*    94948    96551: contig of 1604 bp in length
*    96552    96651: gap of unknown length
*    96652    98298: contig of 1647 bp in length
*    98299    98398: gap of unknown length
*    98399   100654: contig of 2256 bp in length
*   100655   100754: gap of unknown length
*   100755   101948: contig of 1194 bp in length
*   101949   102048: gap of unknown length
*   102049   103515: contig of 1467 bp in length
*   103516   103615: gap of unknown length
*   103616   105211: contig of 1596 bp in length
*   105212   105311: gap of unknown length
*   105312   106435: contig of 1124 bp in length
*   106436   106535: gap of unknown length
*   106536   107744: contig of 1209 bp in length
*   107745   107844: gap of unknown length
*   107845   109563: contig of 1719 bp in length
*   109564   109663: gap of unknown length
*   109664   112050: contig of 2387 bp in length
*   112051   112150: gap of unknown length
*   112151   113412: contig of 1262 bp in length
*   113413   113512: gap of unknown length
*   113513   114769: contig of 1257 bp in length
*   114770   114869: gap of unknown length
*   114870   116333: contig of 1464 bp in length
*   116334   116433: gap of unknown length
*   116434   117929: contig of 1496 bp in length
*   117930   118029: gap of unknown length
*   118030   119642: contig of 1613 bp in length
*   119643   119742: gap of unknown length
*   119743   121418: contig of 1676 bp in length
*   121419   121518: gap of unknown length
*   121519   123313: contig of 1795 bp in length
*   123314   123413: gap of unknown length
*   123414   124693: contig of 1280 bp in length
*   124694   124793: gap of unknown length
*   124794   126213: contig of 1420 bp in length
*   126214   126313: gap of unknown length
*   126314   128131: contig of 1818 bp in length
*   128132   128231: gap of unknown length
*   128232   129941: contig of 1710 bp in length
*   129942   130041: gap of unknown length
*   130042   131544: contig of 1503 bp in length
*   131545   131644: gap of unknown length
*   131645   133140: contig of 1496 bp in length
*   133141   133240: gap of unknown length
*   133241   134536: contig of 1296 bp in length
*   134537   134636: gap of unknown length

alignment_scores:
            Quality:   81.50              Length:    105
              Ratio:  1.509                Gaps:      4
 Percent Similarity:  51.429   Percent Identity: 28.571

alignment_block:
 US-10-048-197-2 x AC094388   ..

 Align seg 1/1  to: AC094388  from: 1  to: 140984

       5 HisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGl 21
         ||||||           |||:::::::|||:::   |||:::|||   ||
   47362 CACATATTGACAAAGAAGTCCGTCATGCTGGCATTGCTTATAGTGCCAGG 47411

      21 y.........CysValSerThrGlyAsnValAlaMetLysGluGln.... 33
         |         |||   |||:::|||   ::::::::::::::::::
   47412 GCATGGCAGGTGTCGATCCTCTGGCTTTCTCTCGCTCCGGCATAGACTAT 47461

      34 .....AsnGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsn 48
            :::   |||:::   |||:::   |||:::||||||      |||
   47462 TTGTTGAATTTCAGGCCAGTGAGAGACCCATTGTCAAAGGGGGAAGAAAC 47511

      49 LysGln.......................................... 50
         :::|||
   47512 TCCCAAGAACACGTGTGGTGGGCCTCCATCTTCCGCACACAGATGCATTA 47561

      51 ..................GluIleSerSerArgPheGly.......... 57
                          ||||||   ||||||||||||
   47562 ACAAGCCACATTATTGGGAGAAATTGTGAGCAGGTTTGGGGAATGGTTAT 47611

      58 ..SerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHis 73
         :::      |||:::   :::|||:::|||:::      |||   :::
   47612 ATGCCCGACTATCGATGTGCATAATGGTTGTGATTCGCGTTTTCCTTTAT 47661

      74 ThrAlaIleLeuAla 78
         ||||||::::::|||
   47662 ACAGCAATGTTTGCC 47676

seq_name: gb_htg:AL627314

seq_documentation_block:
LOCUS       AL627314              196632 bp    DNA     linear   HTG 30-JAN-2002
DEFINITION  Mus musculus chromosome 4 clone RP23-354H24, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL627314
VERSION     AL627314.5  GI:18152585
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Brown,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 15, 2002 this sequence version replaced gi:17977916.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM354H24
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 196587 bases at least Q40
            Consensus quality: 196618 bases at least Q30
            Consensus quality: 196628 bases at least Q20
            Insert size: 196632; sum-of-contigs
            Insert size: 213967; 14.2% error; agarose-fp
            Quality coverage: 14.38x in Q20 bases; sum-of-contigs Quality
            coverage: 13.21x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES              Location/Qualifiers
     source           1. .196632
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /chromosome="4"
                      /clone="RP23-354H24"
                      /clone_lib="RPCI-23"
     misc_feature     1. .196632
                      /note="assembly_fragment:00416"
BASE COUNT    51927 a  47873 c  46075 g  50757 t
ORIGIN


alignment_scores:
            Quality:   81.50            Length:     77
              Ratio:  1.598              Gaps:      1
 Percent Similarity:  66.234   Percent Identity: 27.273

alignment_block:
 US-10-048-197-2 x AL627314   ..

 Align seg 1/1  to: AL627314  from: 1  to: 196632

      31 LysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLysGlyLysTh 47
         |||::::::||||||   ::::::         ::::::   :::   ::
  176506 AAAAAAAAAAACCAAACCTCCATGCCTTTCCCTTTCTTTCTCTCTTGCTC 176555

      47 rAsnLysGlnGluIleSerSerArgPheGlySerAlaAspSerIleSerP 64
         :|||:::   ::::::||||||   |||      :::   |||:::   |
  176556 TAATTCGCCCTCACTCTCTTCCTTCTTTCCTCCCTCTCCCTCTCTTTGTT 176605

      64 heMetIleValValIleLysPheGlyHisThrAlaIleLeuAlaProAsn 80
         ||   :::   :::::::::|||      :::   :::|||:::|||:::
  176606 TCTGTTTGTTTCTTCTTTCCTTTCTCCTTTCTTTCTTCCTTTCTCCTTCT 176655

      81 ArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrpValLysPr 97
                ::::::|||||      :::||||||||||::::::   ||
  176656 .........TCCTTCCTTTCTTCCTCTCTTTCCTTCCTTTTTCTTCCTCC 176696

      97 oTyrArgProLysAsnLeuSerPheTyrLeu 107
         |:::   |||::::::|||      :::|||
```

```
  176697 CTTCCTCCCTTCCTCCCTTCCTCCCTTCCTC 176727

seq_name: gb_htg:AL627253

seq_documentation_block:
LOCUS       AL627253             259837 bp    DNA     linear   HTG 02-NOV-2001
DEFINITION  Mus musculus chromosome 4 clone RP23-53P10, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL627253
VERSION     AL627253.6  GI:16596771
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 2, 2001 this sequence version replaced gi:16596729.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM53P10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 254379 bases at least Q40
            Consensus quality: 255449 bases at least Q30
            Consensus quality: 256413 bases at least Q20
            Insert size: 257937; sum-of-contigs
            Insert size: 239237; 12.2% error; agarose-fp
            Quality coverage: 9.95x in Q20 bases; sum-of-contigs Quality
            coverage: 10.97x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES              Location/Qualifiers
     source           1. .259837
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /chromosome="4"
                      /clone="RP23-53P10"
                      /clone_lib="RPCI-23"
     misc_feature     1. .4624
                      /note="assembly_fragment:02442
                      fragment_chain:1
                      clone_end:SP6
                      vector_side:left"
     misc_feature     4725. .7029
                      /note="assembly_fragment:06607
                      fragment_chain:1"
     misc_feature     7130. .17788
                      /note="assembly_fragment:05762
                      fragment_chain:1"
     misc_feature     17889. .39829
                      /note="assembly_fragment:01743
                      fragment_chain:1"
     misc_feature     39930. .42128
                      /note="assembly_fragment:05183"
     misc_feature     42229. .56352
                      /note="assembly_fragment:05923
                      fragment_chain:2"
     misc_feature     56453. .59852
                      /note="assembly_fragment:04699
```

```
                fragment_chain:2"
     misc_feature    59953. .72074
                /note="assembly_fragment:05167
                fragment_chain:2"
     misc_feature    72175. .97636
                /note="assembly_fragment:06788
                fragment_chain:2"
     misc_feature    97737. .110968
                /note="assembly_fragment:06944
                fragment_chain:2"
     misc_feature    111069. .125967
                /note="assembly_fragment:06767
                fragment_chain:2"
     misc_feature    126068. .132726
                /note="assembly_fragment:06915
                fragment_chain:2"
     misc_feature    132827. .147689
                /note="assembly_fragment:04978
                fragment_chain:2"
     misc_feature    147790. .150330
                /note="assembly_fragment:03806
                fragment_chain:2"
     misc_feature    150431. .166175
                /note="assembly_fragment:01355
                fragment_chain:2"
     misc_feature    166276. .180026
                /note="assembly_fragment:05619
                fragment_chain:2"
     misc_feature    180127. .203936
                /note="assembly_fragment:02077
                fragment_chain:2"
     misc_feature    204037. .242827
                /note="assembly_fragment:05860
                fragment_chain:2"
     misc_feature    242928. .251725
                /note="assembly_fragment:05353
                fragment_chain:2"
     misc_feature    251826. .259837
                /note="assembly_fragment:00313
                fragment_chain:2
                clone_end:T7
                vector_side:right"
BASE COUNT    66786 a  63093 c  63992 g  64014 t   1952 others
ORIGIN


alignment_scores:
            Quality:   81.50              Length:     77
              Ratio:  1.598                Gaps:      1
 Percent Similarity:  66.234   Percent Identity: 27.273

alignment_block:
 US-10-048-197-2 x AL627253/rev   ..

 Align seg 1/1  to reverse of: AL627253  from: 1  to: 259837

      31 LysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLysGlyLysTh 47
         |||::::::|||||   ::::::          ::::::   :::   ::
  148893 AAAAAAAAAAACCAAACCTCCATGCCTTTCCCTTTCTTTCTCTCTTGCTC 148844

      47 rAsnLysGlnGluIleSerSerArgPheGlySerAlaAspSerIleSerP 64
         :|||:::   ::::::||||||   |||      :::   |||:::   |
  148843 TAATTCGCCCTCACTCTCTTCCTTCTTTCCTCCCTCTCCCTCTCTTTGTT 148794

      64 heMetIleValValIleLysPheGlyHisThrAlaIleLeuAlaProAsn 80
         ||   :::   :::::::::|||      :::   :::|||:::|||:::
  148793 TCTGTTTGTTTCTTCTTTCCTTTCTCCTTTCTTTCTTCCTTTCTCCTTCT 148744

      81 ArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrpValLysPr 97
                ::::::||||||      :::||||||||||:::::   ||
  148743 .........TCCTTCCTTTCTTCCTCTCTTTCCTTCCTTTTTCTTCCTCC 148703

      97 oTyrArgProLysAsnLeuSerPheTyrLeu 107
```

```
           |:::   |||:::::::|||      :::|||
   148702 CTTCCTCCCTTCCTCCCTTCCTCCCTTCCTC 148672

seq_name: gb_ba:AF169465

seq_documentation_block:
LOCUS       AF169465               267 bp    DNA     linear   BCT 09-AUG-2000
DEFINITION  Neisseria meningitidis strain Z2491 clone Em029 unknown sequence.
ACCESSION   AF169465
VERSION     AF169465.1  GI:9754673
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 267)
  AUTHORS   Perrin,A., Nassif,X. and Tinsley,C.R.
  TITLE     Identification of regions of the chromosome of Neisseria
            meningitidis and Neisseria gonorrhoeae which are specific to
            pathogenic Neisseriae
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 267)
  AUTHORS   Perrin,A., Nassif,X. and Tinsley,C.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-1999) Necker Medicine Faculty, INSERM U411, 156
            rue de Vaugirard, Paris 75015, France
FEATURES             Location/Qualifiers
     source          1. .267
                     /organism="Neisseria meningitidis"
                     /strain="Z2491"
                     /db_xref="taxon:487"
                     /clone="Em029"
BASE COUNT       71 a     70 c     61 g     65 t
ORIGIN


alignment_scores:
            Quality:   80.50              Length:     71
              Ratio:  1.677                Gaps:      1
 Percent Similarity:  67.606   Percent Identity: 33.803

alignment_block:
 US-10-048-197-2 x AF169465   ..

 Align seg 1/1  to: AF169465  from: 1  to: 267

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||:::   :::|||||||   ::::::||||||   ::::::||||||
      14 CGCATCATCGTTTCGGCTGCGTTCGTTGCGTTGGCATTAGCAGGTTGCGG 63

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
         |||   :::||||||:::::::::::|||:::   |||      :::
      64 CTCAATCAATAATGTAACCGTTTCCGACCAGAAACTTCAGGAA...CGTG 110

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
         ::::::     |||   ||||||      :::|||||:::|||
     111 CCGCGTTTGCCTTGGGCGTCACCAATGCCGTAAAAATCAGCAACCGCAGC 160

      57 GlySerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHi 73
         :::::::::      |||:::|||      :::|||   |||
     161 AATGAAGGCATACGCATCAACTTTACCGCAACTGTGGGTAAGCGCGTGAC 210

      73 sThrAlaIleLeu 77
         :::|||:::|||
     211 CAATGCTATGTTA 223

seq_name: gb_pat:A68918

seq_documentation_block:
LOCUS       A68918                 267 bp    DNA     linear   PAT 06-MAY-1999
DEFINITION  Sequence 89 from Patent WO9802547.
ACCESSION   A68918
VERSION     A68918.1  GI:4759837
```

```
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 267)
  AUTHORS   Nassif,X., Tinsley,C., Achtman,M., Ruelle,J., Vinals,C. and
            Merker,P.
  TITLE     DNA AND SPECIFIC PROTEINS OR PEPTIDES OF THE NEISSERIA MENINGITIDIS
            SPECIES BACTERIA, METHOD FOR OBTAINING THEM AND THEIR BIOLOGICAL
            APPLICATIONS
  JOURNAL   Patent: WO 9802547-A 89 22-JAN-1998;
            INST NAT SANTE RECH MED (FR)
COMMENT     Other publication FR 2751000 19980116.
FEATURES             Location/Qualifiers
     source          1. .267
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT       71 a     70 c     61 g     65 t
ORIGIN


alignment_scores:
            Quality:   80.50             Length:     71
              Ratio:  1.677               Gaps:      1
 Percent Similarity:  67.606   Percent Identity: 33.803

alignment_block:
 US-10-048-197-2 x A68918   ..

 Align seg 1/1  to: A68918  from: 1  to: 267

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||:::   :::||||||   ::::::||||||   ::::::||||||
      14 CGCATCATCGTTTCGGCTGCGTTCGTTGCGTTGGCATTAGCAGGTTGCGG 63

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
          |||   :::||||||:::::::::::::|||:::   |||      :::
      64 CTCAATCAATAATGTAACCGTTTCCGACCAGAAACTTCAGGAA...CGTG 110

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
           ::::::      |||   ||||||      :::|||||:::|||
     111 CCGCGTTTGCCTTGGGCGTCACCAATGCCGTAAAAATCAGCAACCGCAGC 160

      57 GlySerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHi 73
         :::::::::      |||:::|||      :::|||   |||
     161 AATGAAGGCATACGCATCAACTTTACCGCAACTGTGGGTAAGCGCGTGAC 210

      73 sThrAlaIleLeu 77
          :::|||:::|||
     211 CAATGCTATGTTA 223

seq_name: gb_pat:AX024066

seq_documentation_block:
LOCUS       AX024066                339 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 9 from Patent FR2785293.
ACCESSION   AX024066
VERSION     AX024066.1  GI:10184378
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Nassif,X., Tinsley,C., Aujame,L., Perrin,A., Rokbi,B.,
            Bouchardon,A. and Renauld,M.G.
  JOURNAL   Patent: FR 2785293-A 9 05-MAY-2000;
            PASTEUR MERIEUX SERUMS VACC (FR)
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
     CDS             1. .339
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC08909.1"
                     /db_xref="GI:10184379"
                     /translation="MNTRIIVSAAFVALALAGCGSINNVTVSDQKLQERAAFALGVSQ
                     NAVKISNRSNESIRINFTATVGKRVSQCYVTSVISTIGVTTSDAICLGGGTHKGKSQC
                     NALLKAAGRC"
BASE COUNT       93 a     84 c     84 g     78 t
ORIGIN


alignment_scores:
            Quality:   80.00             Length:     89
              Ratio:  1.455               Gaps:      1
 Percent Similarity:  61.798   Percent Identity: 29.213

alignment_block:
 US-10-048-197-2 x AX024066   ..

 Align seg 1/1  to: AX024066  from: 1  to: 339

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||:::   :::||||||   ::::::||||||   ::::::||||||
      10 CGCATCATCGTTTCGGCTGCGTTCGTTGCGTTGGCATTAGCAGGTTGCGG 59

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
          |||   :::||||||:::::::::::::|||:::   |||
      60 CTCAATCAATAATGTAACCGTTTCCGACCAGAAACTTCAGGAACGTGCCG 109

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
           :::      :::   |||      :::||||||:::|||
     110 CGTTTGCCTTGGGCGTCAGCCAAAATGCCGTAAAAATCAGCAACCGCAGC 159

      57 GlySerAlaAspSerIleSerPheMetIleValVal.IleLysPheGlyH 73
         :::::::::      |||:::|||      :::||| ::::::
     160 AATGAAAGCATACGCATCAACTTTACCGCAACTGTGGGTAAGCGCGTGAG 209

      73 isThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIle 89
         :::|||:::|||               |||:::   |||   |||
     210 CCAATGCTATGTTACCAGTGTAATCAGCACAATCGGCGTTACCACTTCCG 259

      90 IleSerPheLeuTrp 94
         :::::::|||:::|||
     260 ATGCAATTTGTTTGG 274

seq_name: gb_pat:AX033695

seq_documentation_block:
LOCUS       AX033695                339 bp    DNA     linear   PAT 21-SEP-2000
DEFINITION  Sequence 1 from Patent WO0044904.
ACCESSION   AX033695
VERSION     AX033695.1  GI:10280379
KEYWORDS    .
SOURCE      Neisseria meningitidis.
  ORGANISM  Neisseria meningitidis
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Thonnard,J.
  TITLE     Basb059 polypeptides from neisseria meningitidis
  JOURNAL   Patent: WO 0044904-A 1 03-AUG-2000;
            SMITHKLINE BEECHAM BIOLOG (BE) ; THONNARD JOELLE (BE)
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Neisseria meningitidis"
                     /db_xref="taxon:487"
BASE COUNT       93 a     84 c     84 g     78 t
ORIGIN


alignment_scores:
            Quality:   80.00             Length:     89
```

```
            Ratio:  1.455                Gaps:      1
 Percent Similarity:  61.798   Percent Identity: 29.213

alignment_block:
 US-10-048-197-2 x AX033695   ..

 Align seg 1/1  to: AX033695  from: 1  to: 339

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||:::   :::||||||   ::::::||||||   ::::::|||||| .
      10 CGCATCATCGTTTCGGCTGCGTTCGTTGCGTTGGCATTAGCAGGTTGCGG 59

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
          |||   :::|||||||:::::::::::::|||:::   |||
      60 CTCAATCAATAATGTAACCGTTTCCGACCAGAAACTTCAGGAACGTGCCG 109

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
                :::     :::   |||      :::|||||||:::|||
     110 CGTTTGCCTTGGGCGTCAGCCAAAATGCCGTAAAAATCAGCAACCGCAGC 159

      57 GlySerAlaAspSerIleSerPheMetIleValVal.IleLysPheGlyH 73
         :::::::::      |||:::|||      :::||| ::::::
     160 AATGAAAGCATACGCATCAACTTTACCGCAACTGTGGGTAAGCGCGTGAG 209

      73 isThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIle 89
         :::|||:::|||                  |||:::   |||   |||
     210 CCAATGCTATGTTACCAGTGTAATCAGCACAATCGGCGTTACCACTTCCG 259

      90 IleSerPheLeuTrp 94
         ::::::|||:::|||
     260 ATGCAATTTGTTTGG 274

seq_name: gb_in:CEB0491

seq_documentation_block:
LOCUS       CEB0491                 38240 bp    DNA     linear   INV 23-JAN-2002
DEFINITION  Caenorhabditis elegans cosmid B0491, complete sequence.
ACCESSION   Z49907
VERSION     Z49907.1  GI:872034
KEYWORDS    HTG; Calcium channel alpha-2 subunit; Chloride channel; Cuticle
            collagen; Diphthine synthasel; Ligand gated ionic channel; SQT-1;
            Transfer RNA; tRNA-Ala.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (sites)
  AUTHORS   none.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  REMARK    The C.elegans Sequencing Consortium.
REFERENCE   2  (bases 1 to 38240)
  AUTHORS   Sulston,J.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUN-1995) Nematode Sequencing Project, Sanger
            Institute, Hinxton, Cambridge CB10 1SA, England and Department of
            Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is not the entire insert of clone B0491.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone B0491 is at 1 in this sequence. The true
            right end of clone B0491 is at 4602 in
            sequence Z48367.
            The true left end of clone C33B4 is at 38141 in this sequence. The
            true right end of clone T24F1 is at 9741 in this sequence. The
            start of this sequence (1. .101) overlaps with the end of sequence
            Z49912.
            The end of this sequence (38141. .38240) overlaps with the start of
            sequence Z48367.
            For a graphical representation of this sequence and its analysis
            see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
            name=B0491
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
FEATURES                Location/Qualifiers
     source             1. .38240
                        /organism="Caenorhabditis elegans"
                        /db_xref="taxon:6239"
                        /chromosome="II"
                        /clone="B0491"
     tRNA               complement(6683. .6754)
                        /gene="B0491.t2"
                        /note="TGC Ala A-tRNA
                        predicted using tRNAscan-SE-1.11
                        preliminary prediction
                        similar to tRNA-Ala"
                        /product="tRNA-Ala"
     gene               complement(6683. .6754)
                        /gene="B0491.t2"
     tRNA               7178. .7249
                        /gene="B0491.t1"
                        /note="TGC Ala A-tRNA
                        predicted using tRNAscan-SE-1.11
                        preliminary prediction
                        similar to tRNA-Ala"
                        /product="tRNA-Ala"
     gene               7178. .7249
                        /gene="B0491.t1"
     gene               complement(join(9398. .10231,10279. .10419))
                        /gene="B0491.2"
     CDS                complement(join(9398. .10231,10279. .10419))
                        /gene="B0491.2"
                        /note="cuticle collagen SQT-1, contains similarity to Pfam
                        domain: PF01391 (Collagen triple helix repeat (20
                        copies)), Score=80.1, E-value=1.5e-20, N=2; PF01484
                        (Nematode cuticle collagen N-terminal domain), Score=33.1,
                        E-value=2e-06, N=1
                        cDNA EST EMBL:Z14426 comes from this gene
                        cDNA EST yk171d9.5 comes from this gene
                        cDNA EST yk171d9.3 comes from this gene
                        cDNA EST yk303b12.3 comes from this gene
                        cDNA EST yk303b12.5 comes from this gene"
                        /codon_start=1
                        /protein_id="CAA90084.1"
                        /db_xref="GI:3873801"
                        /db_xref="SWISS-PROT:P12114"
                        /translation="MSVKLACYVTASVTVATLMVCFMTMSTIYSEVDGFREKLDTEMN
                        VFRQSTNGLWKDIVVIGRSSKRVRRQYEETNATPTPHADGSPSAPPGQPPAVPPVFNQ
                        PKTPNGANGNGPTCNCNADNKCPAGPSGPKGVPGVPGLDGVPGLDGVPGVGADDIAPQ
                        RESVGCFTCPQGPVGPPGALGRPGPRGLPGPRGQNGNPGRDGQPGHPGEQGSSGQIGK
                        IGEPGPPGEKGRDAEHPIGRPGPKGPRGDQGPTGPAGQNGLHGPPGEPGTVGPEGPSG
                        KQGRQGPDGTQGETGPDGRPGKDAEYCQCPDKSPPSEAVNANRGYRNI"
     gene               complement(join(11220. .11324,11372. .11872,12132. .12200,
                        12250. .12252))
                        /gene="B0491.3"
     CDS                complement(join(11220. .11324,11372. .11872,12132. .12200,
                        12250. .12252))
                        /gene="B0491.3"
                        /note="cDNA EST EMBL:M80153 comes from this gene
                        cDNA EST EMBL:M80154 comes from this gene
                        cDNA EST yk144d8.5 comes from this gene
                        cDNA EST yk144d8.3 comes from this gene
```

```
                cDNA EST yk306d2.3 comes from this gene
                cDNA EST yk306d2.5 comes from this gene"
                /codon_start=1
                /protein_id="CAA90085.1"
                /db_xref="GI:3873802"
                /db_xref="SPTREMBL:Q17510"
                /translation="MNFDEIDKTFGTKNRDQGYDLLKARLDKGDRSVEVLWRLAQVIH
                EKSACVPKAQRKASVTEGLKFAEEAVQKDPNHFKALKWNAVLTGQATEYMATKEKLEC
                SKKFKELLDKTLAKEPKDTAVLHLRGRYKYSVASLTWLEKKLAATFYQQPPSHSYEEA
                NEDFLAAYKVNPKWMENTFYVSKCYVAIKDKANARKYLTEVCDIEPYSDAEQEFADDA
                KQMLSKL"
gene            complement(join(12961. .13068,13118. .13252,13294. .13421,
                13939. .14074,14124. .14454,14503. .14622,14737. .14849,
                14896. .14976,15027. .15139,16101. .16200))
                /gene="B0491.4"
CDS             complement(join(12961. .13068,13118. .13252,13294. .13421,
                13939. .14074,14124. .14454,14503. .14622,14737. .14849,
                14896. .14976,15027. .15139,16101. .16200))
                /gene="B0491.4"
                /note="weak similarity with the extracellular domains of
                some ligand gated ionic channels
                cDNA EST yk271a4.5 comes from this gene
                cDNA EST yk271a4.3 comes from this gene"
                /codon_start=1
                /protein_id="CAA90086.1"
                /db_xref="GI:3873803"
                /db_xref="SPTREMBL:Q17511"
                /translation="MSFLVLFGALLLIQTSEQMVQCRYTKNLTNRADGGDEHCLFYYL
                LNKEYEKSGSVHSLSAYPPNSDEMTITIENVLVKYVELVKGSSYQFNVFGDIYLNWKD
                ERMKWDKKGEFENYEHLYIFNSSAIWTPHVIDHTLCSEGGCSYSVDDVDIYDDGTIYA
                RIQFKYLASCTVDYRKFPEEDDSCCIFFTAFEPNVEQTTLVLEGKEKEKMNRPVYAQK
                IYEKEKQLSDIVQEHSPWVVVEKTMSTSNLGGLKSLQVMQVCVRAEKQMSTVRVALLF
                PVTLATYVMLASPLFGDLRTQIYVKLFSLHIQTLCFLYLCTITPLNGFLGVRPRIYIF
                YELIFTISFMSILVTLVVMALSRVKRNVPPSHRLFLSAKLINRVVCCIEPERSDAYHR
                YVEEDTPESRTTSEVSQQQSVIRPGAELDYTQDWRHIYLAMNNMFSALMFSVFLFMTF
                LEFF"
gene            complement(join(16538. .16734,17253. .17480,17538. .17675,
                17735. .17951,18002. .18122,18203. .18285))
                /gene="B0491.5"
CDS             complement(join(16538. .16734,17253. .17480,17538. .17675,
                17735. .17951,18002. .18122,18203. .18285))
                /gene="B0491.5"
                /note="cDNA EST EMBL:M75776 comes from this gene; cDNA EST
                yk56b6.5 comes from this gene
                cDNA EST yk56b6.3 comes from this gene; cDNA EST yk121e5.5
                comes from this gene
                cDNA EST yk114g1.5 comes from this gene; cDNA EST
                yk70e11.5 comes from this gene
                cDNA EST yk187e10.3 comes from this gene; cDNA EST
                yk187e10.5 comes from this gene
                cDNA EST yk143h7.5 comes from this gene; cDNA EST
                yk167a4.5 comes from this gene
                cDNA EST yk501d3.3 comes from this gene; cDNA EST
                yk375c6.3 comes from this gene
                cDNA EST yk375c6.5 comes from this gene; cDNA EST
                yk341h7.3 comes from this gene
                cDNA EST yk341h7.5 comes from this gene; cDNA EST
                yk280b1.3 comes from this gene
                cDNA EST yk280b1.5 comes from this gene; cDNA EST
                yk292g4.3 comes from this gene
                cDNA EST yk292g4.5 comes from this gene; cDNA EST
                yk262e9.3 comes from this gene
                cDNA EST yk262e9.5 comes from this gene; cDNA EST
                yk236b10.3 comes from this gene
                cDNA EST yk236b10.5 comes from this gene; cDNA EST
                yk222e2.3 comes from this gene
                cDNA EST yk222e2.5 comes from this gene; cDNA EST
                yk228g9.5 comes from this gene
                cDNA EST yk290f5.5 comes from this gene; cDNA EST
                yk329h3.5 comes from this gene
                cDNA EST yk344g7.5 comes from this gene; cDNA EST
                yk423a6.5 comes from this gene
                cDNA EST yk452b11.5 comes from this gene; cDNA EST
                yk475b1.5 comes from this gene
                cDNA EST yk483f10.5 comes from this gene; cDNA EST
                yk550e11.3 comes from this gene
                cDNA EST yk573d2.3 comes from this gene; cDNA EST
                yk587h1.3 comes from this gene
                cDNA EST yk598g2.3 comes from this gene; cDNA EST
                yk644a1.3 comes from this gene
                cDNA EST yk497h9.5 comes from this gene; cDNA EST
                yk501d3.5 comes from this gene
                cDNA EST yk502a1.5 comes from this gene; cDNA EST
                yk550e11.5 comes from this gene
                cDNA EST yk573d2.5 comes from this gene; cDNA EST
                yk587h1.5 comes from this gene
                cDNA EST yk598g2.5 comes from this gene; cDNA EST
                yk644a1.5 comes from this gene"
                /codon_start=1

alignment_scores:
            Quality:   79.00            Length:    102
              Ratio:  1.234              Gaps:      6
 Percent Similarity:  62.745   Percent Identity: 30.392

alignment_block:
 US-10-048-197-2 x CEB0491/rev   ..

 Align seg 1/1  to reverse of: CEB0491  from: 1  to: 38240

      10 IleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGl 26
         :::|||:::||||||:::::|||   ::::::   |||:::   :::||
   10632 TTATCCAGTCTTTTATCGTCTCTGCGTCTCGCACACTGTCTC...GTCGG 10586

      26 yAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluGlnThr...I 42
         |||||||      |||:::::::::   :::::::::::::::      :
   10585 TAATGTAAAAAAAAAAACAAAAAAAACTTGAAAACTTGCGGAAAGATTTCC 10536

      42 leIleLysGly...LysThrAsnLysGlnGluIleSerSerArgPheGly 57
         ::||||||||||   ::::::::::::::::::   |||      |||:::
   10535 TAATAAAAGGAGACGAATCCGGTAGGAAATCGACAAGTCTCTGCTTCTCG 10486

      58 SerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHisTh 74
         ::::::      ::::::|||
   10485 AAGACTAGGTGAGTTACATTT..........................  10465

      74 rAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIleS 91
               |||   ||||||   :::   |||||||||:::||||||||||
   10464 .......CTTTGGCCGAAT...TATTTTGAAATCCTAAATTTGATAATTT 10425

      91 erPheLeuTrpValLysProTyrArgProLysAsnLeuSerPheTyrLeu 107
         ||   :::   ||||||                  |||:::   |||:::
   10424 CAGAGATGTCTGTAAAA.................CTTGCGTGTTATGTG 10393

     108 ThrAla 109
         ||||||
   10392 ACGGCT 10387

seq_name: gb_ba:NMA2Z2491

seq_documentation_block:
LOCUS       NMA2Z2491             349061 bp    DNA     linear   BCT 04-DEC-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
            segment 2/7.
ACCESSION   AL162753 AL157959
VERSION     AL162753.2  GI:7379120
KEYWORDS    .
SOURCE      Neisseria meningitidis Z2491.
  ORGANISM  Neisseria meningitidis Z2491
            Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE   1  (bases 1 to 349061)
  AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
            Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
            Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
            Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
```

```
            Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
            Skelton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
  TITLE     Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
  JOURNAL   Nature 404 (6777), 502-506 (2000)
  MEDLINE   20222556
REFERENCE   2  (bases 1 to 349061)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT     Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES             Location/Qualifiers
     source          1. .349061
                     /organism="Neisseria meningitidis Z2491"
                     /strain="Z2491"
                     /db_xref="taxon:122587"
                     /note="serogroup: A"
     repeat_unit     complement(24. .206)
                     /note="ATR repeat; hmmfs hit to HMM ATR (1 - 183), score:
                     310.39"
                     /label=ATR
     RBS             209. .212
     gene            220. .681
                     /gene="NMA0368"
     CDS             220. .681
                     /gene="NMA0368"
                     /note="NMA0368, probable integral membrane protein, len:
                     153 aa; contains four probable transmembrane domains"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="CAB83669.1"
                     /db_xref="GI:7379121"
                     /db_xref="SPTREMBL:Q9JWH7"
                     /translation="MQEQNRKSSFPIVMLLVSVALWIASLSNVAFYLGNHGSMEGLTV
                     LILGSIFASLDIRYCAVYANYVWLAAIVLLALRKKVVPVHAAFWGLALVAFSVKAVYV
                     DEAGNTSDIVRYGAGFYLWYAAFAVASIGTFAGKNKERKAASAADGTKNDV"
     gene            671. .1492
                     /gene="hemK"
     CDS             671. .1492
                     /gene="hemK"
                     /note="NMA0369, hemK, HemK protein, len: 273 aa; similar
                     to e.g. HEMK_ECOLI P37186 HEMK protein (277 aa), fasta
                     scores; E(): 0, 42.3% identity in 279 aa overlap. Contains
                     PS00092 N-6 Adenine-specific DNA methylases signature"
                     /codon_start=1
                     /transl_table=11
                     /product="HemK protein"
                     /protein_id="CAB83670.1"
                     /db_xref="GI:7379122"
                     /db_xref="SPTREMBL:Q9JWH6"
                     /translation="MTFDKWLGLSKLPKNEARMLLQYVSEYTRVQLLTRGGEEMPDEV
                     RQRADRLAQRRLNGEPVAYILGAREFYGRRFTVNPSVLIPRPETEHLVEAVLARLPEN
                     GRVWDLGTGSGAVAVTVALERPDAFVRASDISPPALETARKNAADLGARVEFAYGSWF
                     DTDMPSEGKWDIIVSNPPYIENGDKHLSQGDLRFEPQIALTDFSDGLSCIRTLAQGAP
                     DRLAEGGFLLLEHGFDQGAAVRGVLAENGFSGVETLPDLAGLDRVTLGKYMKHLK"
     misc_feature    831. .840
                     /gene="hemK"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    1162. .1171
                     /gene="hemK"
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    1187. .1207
                     /gene="hemK"
                     /note="PS00092 N-6 Adenine-specific DNA methylases
                     signature"
     misc_feature    complement(1279. .1288)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     RBS             1565. .1568
     gene            1578. .2966
                     /gene="NMA0370"
     CDS             1578. .2966
                     /gene="NMA0370"
                     /note="NMA0370, probable integral membrane protein, len:
                     462 aa; similar to hypothetical proteins e.g. Y325_HAEIN
                     P44640 hypothetical protein HI0325 (450 aa), fasta scores;
                     E(): 0, 49.8% identity in 464 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="CAB83671.1"
                     /db_xref="GI:7379123"
                     /db_xref="SPTREMBL:Q9JWH5"
                     /translation="MNAVVVAVIVMLVLSLSRVHVVLSLTIGAFVGGAVAGMPLQNIA
                     DAAGQVSQAGIIPVFNKGLEGGAKIALSYAMLGAFAMAITHSGLPQQLAGAVVRKLNR
                     GGMPDSVRSGEGVVKWLLLSIILVMGMMSQNIIPIHIAFIPMIVPPLLLVFNRLKIDR
                     RLIACVITFGLVTTYMFLPYGFGAIFLNEILLGNIHSAAPQLDVKNINVMAAMAIPAL
                     GMLAGLLLAFVHYRKPRLYQSNNADTAGNADAANRPQPSAYRSLAAAVAIAVCFAIQL
                     MYEDSLVLGAMLGFAVFMMLGVINRDKANDVFGEGIKMMAMVGFIMIAAQGFAAVMNA
                     TGHIQPLVESSMAIFGNSKGMAALAMLVVGLLVTMGIGSSFSTLPIIAAIYVPLCVGL
                     GFSPLATVAIVGTAGALGDAGSPASDSTLGPTMGLNADGQHDHIRDSVIPTFIHYNIP
                     LLIAGWIAAMVL"
     gene            2967. .3191
                     /gene="slyX"
     CDS             2967. .3191
                     /gene="slyX"
                     /note="NMA0371, slyX, SLYX protein homolog, len: 74 aa;
                     simlar to SLYX_HAEIN P44759 SLYX protein homolog (73 aa),
                     fasta scores; E(): 0.0017, 33.8% identity in 74 aa
                     overlap, and SLYX_ECOLI P30857 SLYX protein (72 aa), fasta
                     scores; E(): 0.24, 32.4% identity in 68 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="SLYX protein homolog"
                     /protein_id="CAB83672.1"
                     /db_xref="GI:7379124"
                     /db_xref="SPTREMBL:Q9JWH4"
                     /translation="MDAVQEFEHRITELEIQSALQEDVIAGLNAMVAELRQTLDLQQA
                     QLRLLYQKMQDRNPDAQEPYSLRDEIPPHY"
     gene            complement(3271. .3423)
                     /gene="NMA0372"
     CDS             complement(3271. .3423)
                     /gene="NMA0372"
                     /note="NMA0372, unknown, questionable CDS, len: 50 aa"
                     /codon_start=1
                     /transl_table=11
                     /product="very hypothetical protein NMA0372"
                     /protein_id="CAB83673.1"
                     /db_xref="GI:7379125"
                     /db_xref="SPTREMBL:Q9JWH3"
                     /translation="MFGRLSPLGSDGFFLAVYEYPVGFICLQGKTAYFRSGGKRFHRI
                     KGILSD"
     misc_feature    3562. .3571
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            complement(3586. .4356)
                     /gene="thiF"
     CDS             complement(3586. .4356)
                     /gene="thiF"
                     /note="NMA0373, thiF, probable ThiF protein, len: 256 aa;
                     similar to e.g. THIF_ECOLI P30138 THIF protein (251 aa),
                     fasta scores; E(): 0, 43.1% identity in 246 aa overlap,
                     and MOEB_ECOLI P12282 molybdopterin biosynthesis MOEB
                     protein. (249 aa), fasta scores; E(): 0, 43.9% identity in
                     244 aa overlap (note that N.m. does not have orthologs of
                     any other molybdopterin biosynthesis proteins). Contains
                     Pfam match to entry PF00899 ThiF_family, ThiF family"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="ThiF protein"
                     /protein_id="CAB83674.1"
                     /db_xref="GI:7379126"
                     /db_xref="SPTREMBL:Q9JWH2"
                     /translation="MTTTEHDNDDAFLLRYSRHILLDEIGIEGQQKLSAAHILVVGCG
                     GLGAAALPYLAASGIGTLTIADSDTVELHNLQRQVAFDEGDVGKLKTEALADRLRHIN
                     HTVDVRTINEKLDGCRLTGLVQAADIVLDCCDNYATRQAVNRACVQAKTPLVSGAAVR
                     FEGQLAVYRPDLPDSPCYACLFDGGSASDGICSLFGVFSPLVGIIGSTQAAEALKILL
                     DAGEPSHGRLAVYRALEGGWQYFDLPRNPECPVCGAER"
     misc_feature    3787. .3796
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     misc_feature    complement(3847. .4257)
                     /gene="thiF"
                     /note="Pfam match to entry PF00899 ThiF_family, ThiF
                     family, score 186.60, E-value 4.1e-52"
     misc_feature    complement(4418. .4427)
                     /note="Core DNA uptake sequence: gccgtctgaa"
                     /label=DUS
     gene            4471. .7224
                     /gene="ppc"
     CDS             4471. .7224
                     /gene="ppc"
                     /EC_number="4.1.1.31"
                     /note="NMA0374, ppc, phosphoenolpyruvate carboxylase, len:
                     917 aa; simlar to many e.g. CAPP_RHOPA O32483
                     phosphoenolpyruvate carboxylase (EC 4.1.1.31) (936 aa),
                     fasta scores; E(): 0, 43.3% identity in 928 aa overlap.
                     Contains 2x Pfam match to entry PF00311 PEPcase,
                     Phosphoenolpyruvate carboxylase, PS00017 ATP/GTP-binding
                     site motif A (P-loop), and PS00393 Phosphoenolpyruvate
                     carboxylase active site 2"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoenolpyruvate carboxylase"
                     /protein_id="CAB83675.1"
                     /db_xref="GI:7379127"
                     /db_xref="SPTREMBL:Q9JWH1"

alignment_scores:
           Quality:   79.00            Length:     89
             Ratio:  1.436              Gaps:      1
 Percent Similarity:  61.798   Percent Identity: 29.213

alignment_block:
 US-10-048-197-2 x NMA222491   ..

 Align seg 1/1  to: NMA222491  from: 1  to: 349061

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||:::   :::||||||   ::::::||||||   ::::::||||||
  149160 CGCATCATCGTTTCGGCTGCGTTCGTTGCGTTGGCATTAGCAGGTTGCGG 149209

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
          |||   :::||||||:::::::::::||:::   |||
  149210 CTCAATCAATAATGTAACCGTTTCCGACCAGAAACTTCAGGAACGTGCCG 149259

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
                :::     :::   |||      :::||||||:::|||
  149260 CGTTTGCCTTGGGCGTCAGCCCAAATGCCGTAAAAATCAGCAACCGCAGC 149309

      57 GlySerAlaAspSerIleSerPheMetIleValVal.IleLysPheGlyH 73
         :::::::::      |||:::|||      :::||| ::::::
  149310 AATGAAGGCATACGCATCAACTTTACCGCAACTGTGGGTAAGCGCGTGAG 149359

      73 isThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIle 89
           :::|||:::|||               |||:::   |||   |||
  149360 CCAATGCTATGTTACCAGTGTAATCAGCACAATCGGCGTTACCACTTCCG 149409

      90 IleSerPheLeuTrp 94
         ::::::|||:::|||
  149410 ATGCAATTTGTTTGG 149424

seq_name: gb_htg:AC102510

seq_documentation_block:
LOCUS       AC102510              65312 bp    DNA     linear   HTG 23-NOV-2001
DEFINITION  Mus musculus clone RP24-346C14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC102510
VERSION     AC102510.1  GI:17061596
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 65312)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-346C14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 65312)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L18939
                Center clone name: 346_C_14
            --------------.
            * NOTE: This record contains 83 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      661: contig of 661 bp in length
            *      662 761: gap of      100 bp
            *      762     1435: contig of 674 bp in length
            *     1436 1535: gap of      100 bp
            *     1536     2216: contig of 681 bp in length
            *     2217 2316: gap of      100 bp
            *     2317     3017: contig of 701 bp in length
```

```
*     3018 3117: gap of      100 bp
*     3118     3807: contig of 690 bp in length
*     3808 3907: gap of      100 bp
*     3908     4576: contig of 669 bp in length
*     4577 4676: gap of      100 bp
*     4677     5353: contig of 677 bp in length
*     5354 5453: gap of      100 bp
*     5454     6134: contig of 681 bp in length
*     6135 6234: gap of      100 bp
*     6235     6904: contig of 670 bp in length
*     6905 7004: gap of      100 bp
*     7005     7690: contig of 686 bp in length
*     7691 7790: gap of      100 bp
*     7791     8484: contig of 694 bp in length
*     8485 8584: gap of      100 bp
*     8585     9275: contig of 691 bp in length
*     9276 9375: gap of      100 bp
*     9376    10095: contig of 720 bp in length
*    10096 10195: gap of      100 bp
*    10196    10912: contig of 717 bp in length
*    10913 11012: gap of      100 bp
*    11013    11713: contig of 701 bp in length
*    11714 11813: gap of      100 bp
*    11814    12507: contig of 694 bp in length
*    12508 12607: gap of      100 bp
*    12608    13304: contig of 697 bp in length
*    13305 13404: gap of      100 bp
*    13405    14076: contig of 672 bp in length
*    14077 14176: gap of      100 bp
*    14177    14866: contig of 690 bp in length
*    14867 14966: gap of      100 bp
*    14967    15644: contig of 678 bp in length
*    15645 15744: gap of      100 bp
*    15745    16418: contig of 674 bp in length
*    16419 16518: gap of      100 bp
*    16519    17209: contig of 691 bp in length
*    17210 17309: gap of      100 bp
*    17310    18000: contig of 691 bp in length
*    18001 18100: gap of      100 bp
*    18101    18794: contig of 694 bp in length
*    18795 18894: gap of      100 bp
*    18895    19595: contig of 701 bp in length
*    19596 19695: gap of      100 bp
*    19696    20352: contig of 657 bp in length
*    20353 20452: gap of      100 bp
*    20453    21143: contig of 691 bp in length
*    21144 21243: gap of      100 bp
*    21244    21907: contig of 664 bp in length
*    21908 22007: gap of      100 bp
*    22008    22700: contig of 693 bp in length
*    22701 22800: gap of      100 bp
*    22801    23489: contig of 689 bp in length
*    23490 23589: gap of      100 bp
*    23590    24279: contig of 690 bp in length
*    24280 24379: gap of      100 bp
*    24380    25063: contig of 684 bp in length
*    25064 25163: gap of      100 bp
*    25164    25862: contig of 699 bp in length
*    25863 25962: gap of      100 bp
*    25963    26655: contig of 693 bp in length
*    26656 26755: gap of      100 bp
*    26756    27442: contig of 687 bp in length
*    27443 27542: gap of      100 bp
*    27543    28228: contig of 686 bp in length
*    28229 28328: gap of      100 bp
*    28329    29018: contig of 690 bp in length
*    29019 29118: gap of      100 bp
*    29119    29808: contig of 690 bp in length
*    29809 29908: gap of      100 bp
*    29909    30602: contig of 694 bp in length
*    30603 30702: gap of      100 bp
*    30703    31398: contig of 696 bp in length
*    31399 31498: gap of      100 bp
*    31499    32195: contig of 697 bp in length
*    32196 32295: gap of      100 bp
*    32296    33012: contig of 717 bp in length
*    33013 33112: gap of      100 bp
*    33113    33824: contig of 712 bp in length
*    33825 33924: gap of      100 bp
*    33925    34613: contig of 689 bp in length
*    34614 34713: gap of      100 bp
*    34714    35402: contig of 689 bp in length
*    35403 35502: gap of      100 bp
*    35503    36190: contig of 688 bp in length
*    36191 36290: gap of      100 bp
*    36291    36951: contig of 661 bp in length
*    36952 37051: gap of      100 bp
*    37052    37741: contig of 690 bp in length
*    37742 37841: gap of      100 bp
*    37842    38530: contig of 689 bp in length
*    38531 38630: gap of      100 bp
*    38631    39357: contig of 727 bp in length
*    39358 39457: gap of      100 bp
*    39458    40125: contig of 668 bp in length
*    40126 40225: gap of      100 bp
*    40226    40920: contig of 695 bp in length
*    40921 41020: gap of      100 bp
*    41021    41714: contig of 694 bp in length
*    41715 41814: gap of      100 bp
*    41815    42519: contig of 705 bp in length
*    42520 42619: gap of      100 bp
*    42620    43305: contig of 686 bp in length
*    43306 43405: gap of      100 bp
*    43406    44103: contig of 698 bp in length
*    44104 44203: gap of      100 bp
*    44204    44876: contig of 673 bp in length
*    44877 44976: gap of      100 bp
*    44977    45644: contig of 668 bp in length
*    45645 45744: gap of      100 bp
*    45745    46429: contig of 685 bp in length
*    46430 46529: gap of      100 bp
*    46530    47219: contig of 690 bp in length
*    47220 47319: gap of      100 bp
*    47320    47994: contig of 675 bp in length
*    47995 48094: gap of      100 bp
*    48095    48798: contig of 704 bp in length
*    48799 48898: gap of      100 bp
*    48899    49595: contig of 697 bp in length
*    49596 49695: gap of      100 bp
*    49696    50400: contig of 705 bp in length
*    50401 50500: gap of      100 bp
*    50501    51167: contig of 667 bp in length
*    51168 51267: gap of      100 bp
*    51268    51970: contig of 703 bp in length
*    51971 52070: gap of      100 bp
*    52071    52736: contig of 666 bp in length
*    52737 52836: gap of      100 bp
*    52837    53511: contig of 675 bp in length
*    53512 53611: gap of      100 bp
*    53612    54295: contig of 684 bp in length
*    54296 54395: gap of      100 bp

alignment_scores:
            Quality:   78.50              Length:     99
              Ratio:  1.670                Gaps:      3
 Percent Similarity:  47.475   Percent Identity: 27.273

alignment_block:
 US-10-048-197-2 x AC102510   ..

 Align seg 1/1  to: AC102510  from: 1  to: 65312

      8 LeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSe 24
        |||:::|||      |||   |||||||||           ||||||::
   2340 CTGGTCATCTTCAACTTGTCAACTGCTTTGTACTATGAGCAATGTGTAAC 2389
```

```
      24 rThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluGlnT 41
         :   |||:::                              |||:::::::|||
    2390 ACCTGGACAT.........................ACGATGAAGCAGG 2412

      41 hrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPheGly 57
         :::::::::||||||
    2413 ATGTAGTGTCAGGCAAA.............................. 2429

      58 SerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHisTh 74
                                  :::||||||:::   :::::::
    2430 .....................GCTGTCATTAGTGAAGCCAACTC 2452

      74 rAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIleS 91
         :|||:::   :::   :::|||||||||   |||            ||||
    2453 AGCTTTCAAAGTTGTTTCACGTTGGCAAATCATACCCCCAAGCAGTATTT 2502

      91 erPheLeuTrpValLysProTyrArgProLysAsnLeuSerPheTyr 106
         ||   :::|||         :::   :::||||||:::||||||
    2503 CTGTGTTTTGG...........CAAAGTAATAACCTTGAATTCTAT 2537

seq_name: gb_pat:BD003755

seq_documentation_block:
LOCUS       BD003755              8136 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION   BD003755
VERSION     BD003755.1  GI:18631716
KEYWORDS    JP 2001501833-A/75.
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 8136)
  AUTHORS   Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C.,
            Fannon,M. and Dougherty,B.A.
  TITLE     Polynucleotide of Streptococcus pneumoniae and sequence
  JOURNAL   Patent: JP 2001501833-A 75 13-FEB-2001;
            HUMAN GENOME SCIENCES INC
COMMENT     OS   Unidentified
            PN   JP 2001501833-A/75
            PD   13-FEB-2001
            PF   30-OCT-1997 JP 1998520718
            PR   31-OCT-1996 US    60/029960
            PI   CHARLES A KUNSCH,GIL H CHOI,PATRICK J DILLON,CRAIG A ROSEN, PI
              STEVEN C BARASH,
            PI   MICHAEL FANNON,BRIAN A DOUGHERTY
            PC   C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19,
            PC   C12N1/21,
            PC   C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC
            G06F15/40
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   source          1. .8136
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .8136
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT     2249 a   1481 c   1983 g   2423 t
ORIGIN


alignment_scores:
            Quality:   78.00             Length:    127
              Ratio:  1.200               Gaps:      6
 Percent Similarity:  51.181   Percent Identity: 28.346

alignment_block:
 US-10-048-197-2 x BD003755   ..

 Align seg 1/1  to: BD003755  from: 1  to: 8136

      10 IleSerAlaLeuLeuThrAla...LeuLeuValThrGlyCysValSerTh 25
         ::::::||||||:::||||||   :::|||||||||||   |||   ::
    1926 CTAGGAGCCTTGGTAACAGCTGTGATTCTCGTAACGGGCTCTGTTCTAGT 1975

      25 r......GlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
         :       ||||||:::         :::   |||   :::   ::::
    1976 CATTTTGGAAAATGTCACGAAGATTTTGCATCCGCAACCAGTCAATGATG 2025

      40 lnThrIleIle..................................... 43
         ::   |||:::
    2026 AGGGGATTCTCTGGTTAGGAATTATTGCGATTACTATCAATCTGTTAGCG 2075

      44 ..............LysGlyLysThrAsnLysGlnGluIleSerSerAr 55
                       ||||||||||||:::::::::::::|||   |||
    2076 AGTCTGGTGGTTGGTAAGGGAAAGACAAAGAATGAGTCTATTCTGAGTCT 2125

      55 gPheGlySerAlaAspSerIleSerPhe..............MetIleV 67
                     |||:::::::::::::               |||   :
    2126 GCATTTTCTGGAAGATACGCTAGGGTGGGTAGCTGTTATCCTGATGGCGA 2175

      67 alValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 83
         ::|||::::::|||              ||||||   |||
    2176 TTGTTCTTCGATTTACGGACTGGTATATCCTAGATCCT............ 2213

      84 GluIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPr 100
            :::||||||||||:::|||||||||:::   ::::::         ||
    2214 ...CTTTTGTCCCTTGTCATTTCTTTCTTTATTCTTTCAAAAGCCCTTCC 2260

     100 oLysAsnLeuSerPheTyrLeuThrAlaLys 110
         |:::         |||:::   |||   |||
    2261 ACGT.........TTTTGGTCTACACTCAAG 2282

seq_name: gb_ba:AE008532

seq_documentation_block:
LOCUS       AE008532              10393 bp    DNA     linear   BCT 13-SEP-2001
DEFINITION  Streptococcus pneumoniae R6 section 148 of 184 of the complete
            genome.
ACCESSION   AE008532 AE007317
VERSION     AE008532.1  GI:15459334
KEYWORDS    .
SOURCE      Streptococcus pneumoniae R6.
  ORGANISM  Streptococcus pneumoniae R6
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 10393)
  AUTHORS   Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszczak,L., Burgett,S.,
            DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
            Gilmour,R., Glass,J.S., Khoja,H., Kraft,A., LaGace,R.,
            LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
            McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
            Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
            Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
            Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
            and Glass,J.I.
  TITLE     Genome of the Bacterium Streptococcus pneumoniae Strain R6
  JOURNAL   J. Bacteriol. 183 (19), 5709-5717 (2001)
  MEDLINE   21429245
   PUBMED   11544234
REFERENCE   2  (bases 1 to 10393)
  AUTHORS   Hoskins,J.A., Alborn,W. Jr., Arnold,J., Blaszczak,L., Burgett,S.,
            DeHoff,B.S., Estrem,S., Fritz,L., Fu,D.-J., Fuller,W., Geringer,C.,
            Gilmour,R., Glass,J.S., Hann,A., Khoja,H., Kraft,A., LaGace,R.,
            LeBlanc,D.J., Lee,L.N., Lefkowitz,E.J., Lu,J., Matsushima,P.,
            McAhren,S., McHenney,M., McLeaster,K., Mundy,C., Nicas,T.I.,
            Norris,F.H., O'Gara,M., Peery,R., Robertson,G.T., Rockey,P.,
            Sun,P.-M., Winkler,M.E., Yang,Y., Young-Bellido,M., Zhao,G.,
            Zook,C., Baltz,R.H., Jaskunas,S.R., Rosteck,P.R. Jr., Skatrud,P.L.
            and Glass,J.I.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and
            Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
FEATURES             Location/Qualifiers
     source          1. .10393
```

```
                /organism="Streptococcus pneumoniae R6"
                /strain="R6"
                /db_xref="taxon:171101"
     gene       102. .683
                /gene="xpt"
                /note="spr1662"
     CDS        102. .683
                /gene="xpt"
                /EC_number="2.4.2.-"
                /note="This sequence is longer than the other published
                sequence"
                /codon_start=1
                /transl_table=11
                /label=spr1662
                /product="Xanthine phosphoribosyltransferase"
                /protein_id="AAL00465.1"
                /db_xref="GI:15459335"
                /translation="MKLLEERILKDGHILDDNILKVDSFLTHQVDFSLMREIGKVFAE
                KFAATGITKVVTIEASGIAPAVFTAEALNVPMIFAKKAKNITMNEGILTAQVYSFTKQ
                VTSTVSIAGKFLSPEDKVLIIDDFLANGQAAKGLIQIIEQAGATVQAIGIVIEKSFQD
                GRDLLEKAGYPVLSLARLDRFENGQVVFKEADL"
     gene       683. .1945
                /gene="pbuX"
                /note="spr1663"
     CDS        683. .1945
                /gene="pbuX"
                /codon_start=1
                /transl_table=11
                /label=spr1663
                /product="Nucleobase:cation symporter for xanthine"
                /protein_id="AAL00466.1"
                /db_xref="GI:15459336"
                /translation="MQTQEKHSQAAVLGLQHLLAMYSGSILVPIMIATALGYSAEQLT
                YLISTDIFMCGVATFLQLQLNKYFGIGLPVVLGVAFQSVAPLIMIGQSHGSGAMFGAL
                IASGIYVVLVSGIFSKVANLFPSIVTGSVITTIGLTLIPVAIGNMGNNVPEPTGQSLL
                LAAITVLIILLINIFTKGFIKSISILIGLVVGTAIAATMGLVDFSPVAVAPLVHVPTP
                LYFGMPTFEISSIVMMCIIATVSMVESTGVYLALSDITKDPIDSTRLRNGYRAEGLAV
                LLGGIFNTFPYTGFSQNVGLVKLSGIKKRLPIYYAAGFLVLLGLLPKFGALAQIIPSS
                VLGGAMLVMFGFVSIQGMQILARVDFANNEHNFLIAAVSIAAGVGLNNSNLFVSMPTA
                FQMFFSNGIVVSSLLAIVLNAVLNHKKK"
     gene       complement(1996. .2457)
                /gene="dpnD"
     CDS        /note="spr1664"
                complement(1996. .2457)
                /gene="dpnD"
                /EC_number="2.1.1.-"
                /codon_start=1
                /transl_table=11
                /label=spr1664
                /product="Restriction system of S. pneumoniae"
                /protein_id="AAL00467.1"
                /db_xref="GI:15459337"
                /translation="MKTKQLVASEEVYDFLKVIWPDYETESRYDNLSLIVCTLSDPDC
                VRWLSENMKFGDEKQLALMKEKYGWEVGDKLPEWLHSSYHRLLLIGELLESNLKLKKY
                TVEITETLSRLVSIEAENPDEAERLVREKYKSCEIVLDADDFQDYDTSIYE"
     gene       complement(2454. .3218)
                /gene="dpnC"
     CDS        /note="spr1665"
                complement(2454. .3218)
                /gene="dpnC"
                /EC_number="3.1.21.4"
                /codon_start=1
                /transl_table=11
                /label=spr1665
                /product="Type II restriction enzyme DpnI (dpnC)"
                /protein_id="AAL00468.1"
                /db_xref="GI:15459338"
                /translation="MELHFNLELVETYKSNSQKARILTEDWVYRQSYCPNCGNNPLNH
                FENNRPVADFYCNHCSEEFELKSKKGNFSSTINDGAYATMMKRVQADNNPNFFFLTYT
                KNFEVNNFLVLPKQFVTPKSIIQRKPLAPTARRAGWIGCNIDLSQVPSKGRIFLVQDG
                QVRDPEKVTKEFKQGLFLRKSSLSSRGWTIEILNCIDKIEGSEFTLEDMYRFESDLKN
                IFVKNNHIKEKIRQQLQILRDKEIIEFKGRGKYRKL"
     gene       complement(3399. .3824)
                /gene="spr1666"
     CDS        complement(3399. .3824)
                /gene="spr1666"
                /note="Similar to E. coli PaaI protein involved in
                phenylacetic acid degradation"
                /codon_start=1
                /transl_table=11
                /label=spr1666
                /product="Conserved hypothetical protein"
                /protein_id="AAL00469.1"
                /db_xref="GI:15459339"
                /translation="MKYQEETMKDFHFDAISAFENYEIEKMRDGHVVVTTKVVNSSLN
                YYGNAHGGYLFTLCDQISGLVVISLGLDGVTLQSSINYLKAGKLDDVLTIKGECVHQG
                RTTCVMDVDITNQEGRNVCKATFTMFVTGQRSEERRVSI"
     gene       complement(3873. .5354)
                /gene="galT"
     CDS        /note="spr1667"
                complement(3873. .5354)
                /gene="galT"
                /EC_number="2.7.7.10"
                /codon_start=1
                /transl_table=11
                /label=spr1667
                /product="Galactose-1-phosphate uridylyltransferase"
                /protein_id="AAL00470.1"
                /db_xref="GI:15459340"
                /translation="MTLVDKFVTHVISESSFEEMDRIYLTNRVLSRVGEGVLEVETNL
                DKLIDLKDQLVEETVRLETIEDSQTAREILGTELMDLVTPCPSQVNRDFWEAYAYSPE
                QAIEDFYQLSQKNDYIKLKAIAKNIAYRVPSDYGELEITINLSKPEKDPKEIAVAKLV
                QASNYPQCQLCLENEGYHGRVNHPARSNHRIIRFEMVGQEWGFQYSPYAYFNEHCIFL
                DGQHRPMAISRQSFERLLAIVEQFPGYFAGSNADLPIVGGSILTHDHYQGGRHVFPME
                LAPLQKTFRFAGFEQVKAGIIKWPMSVLRLTSDSKEDLINLADKIFQEWRQYSDSSVQ
                ILAETDGTPHHTITPIARKRDGQFELDLVLRDNQTSAEHPDGIYHPHKDVQHIKKENI
                GLIEVMGLAILPPRLKEEVEQVASYLVGEAVTVADYHQEWADQLKSQHPDLTDKEKAL
                AIVKDSVGAIFVRVLEDAGVYKQTEQGQTAFMRFVEQVGILLD"
     gene       complement(5373. .6551)
                /gene="galK"
     CDS        /note="spr1668"
                complement(5373. .6551)
                /gene="galK"
                /EC_number="2.7.1.6"
                /codon_start=1
                /transl_table=11
                /label=spr1668
                /product="Galactokinase"
                /protein_id="AAL00471.1"
                /db_xref="GI:15459341"
                /translation="MTQHLTAETLRKDFLAVFGQEVDQTFFSPGRINLIGEHTDYNGG
                HVFPVAISLGTYGAARKRDDQVLRFYSANFEDKGIIEVPLADLKFEKEHNWTNYPKGV
                LHFLQEAGHVIDKGFDFYVYGNIPNGSGLSSSASLEILTGVVAEHLFDLKLERLDLVK
                IGKQTENNFIGVNSGIMDQFAIGMGADQRAIYLDTNTLEYDLVPLDLKDNVVVIMNTN
                KRRELADSKYNERRAECEKAVEELQVALDIQTLGELDEWAVDQYSYLIKDENRLKRAR
                HAVLENQRTLKAQAALQAGDLETFGRLMNASHVSLEHDYEVTGLELDTLVHTAWAQEG
                VLGARMTGAGFSGCAIALVQKDTVEAFKEAVGKHYEEVVGYAPSFYIAEVAGGTRVLD
                "
     gene       6715. .7722
                /gene="galR"
     CDS        /note="spr1669"
                6715. .7722
                /gene="galR"
                /codon_start=1
                /transl_table=11
                /label=spr1669
                /product="GalR, member of GalR-LacI family of
                transcriptional regulators, binds DNA; regulator of gal
                operon"
                /protein_id="AAL00472.1"
                /db_xref="GI:15459342"
                /translation="MATLKDIAQLASVSIATVSRVLNRDQSLSVTEETRHRILTVAEE
                LGYTKHLKTGESHKPKQKIAIIQWVSEQGELDDLYYYQIRLGIEKRAQELDYDILRYF
                NDHPFTLSEEVIGILCIGKFSRAQISAFEEYQKPLVFLDSDTLSLGHTCIITDFYTAM
                KQVVDYFLSQGMDRIGILTGLEETTDQEEIIQDKRLENFKNYSQARGIYHDELVFQGR
                FTAQSGYDLMKEAIQSLGDQLPPAFFAASDSLAIGALRALQEAGISLPDRVSLISFND
```

```
                    TSLTKQVYPPLSSITVYTEEMGRAGMDILNKEVLHGRKIPSLTMLGTRLTLRESTLNQ
                    E"
     gene           complement(7768. .8805)
                    /gene="adhB"
                    /note="spr1670"
     CDS            complement(7768. .8805)
                    /gene="adhB"
                    /EC_number="1.1.1.1"
                    /note="Putative"
                    /codon_start=1
                    /transl_table=11
                    /label=spr1670
                    /product="Alcohol dehydrogenase"
                    /protein_id="AAL00473.1"
                    /db_xref="GI:15459343"

alignment_scores:
           Quality:   78.00            Length:    127
             Ratio:  1.200              Gaps:      6
 Percent Similarity:  51.181   Percent Identity: 28.346

alignment_block:
 US-10-048-197-2 x AE008532/rev   ..

 Align seg 1/1  to reverse of: AE008532  from: 1  to: 10393

      10 IleSerAlaLeuLeuThrAla...LeuLeuValThrGlyCysValSerTh 25
         ::::::||||||:::||||||   :::|||||||||||||   |||   ::
   10071 CTAGGAGCCTTGGTAACAGCTGTGATTCTCGTAACGGGCTCTGTTCTAGT 10022

      25 r......GlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
         :       ||||||:::         :::   |||   :::   ::::
   10021 CATTTTGGAAAATGTCACGAAGATTTTGCATCCGCAACCAGTCAATGATG 9972

      40 lnThrIleIle..................................... 43
         ::   |||:::
    9971 AGGGGATTCTCTGGTTAGGAATTATTGCGATTACTATCAATCTGTTAGCG 9922

      44 ..............LysGlyLysThrAsnLysGlnGluIleSerSerAr 55
                       ||||||||||||:::::::::::::|||   |||
    9921 AGTCTGGTGGTTGGTAAGGGAAAGACAAAGAATGAGTCTATTCTGAGTCT 9872

      55 gPheGlySerAlaAspSerIleSerPhe.............MetIleV 67
                     |||:::::::::::::              |||   :
    9871 GCATTTTCTGGAAGATACGCTAGGGTGGGTAGCTGTTATCCTGATGGCGA 9822

      67 alValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 83
         ::|||::::::|||           ||||||   |||
    9821 TTGTTCTTCGATTTACGGACTGGTATATCCTAGATCCT............ 9784

      84 GluIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPr 100
            :::||||||||||:::|||||||||:::   ::::::         ||
    9783 ...CTTTTGTCCCTTGTCATTTCTTTCTTTATTCTTTCAAAAGCCCTTCC 9737

     100 oLysAsnLeuSerPheTyrLeuThrAlaLys 110
         |:::         |||:::   |||   |||
    9736 ACGT.........TTTTGGTCTACACTCAAG 9715

seq_name: gb_ba:AE007477

seq_documentation_block:
LOCUS       AE007477               12841 bp    DNA     linear   BCT 31-AUG-2001
DEFINITION  Streptococcus pneumoniae TIGR4  section 160 of 194 of the complete
            genome.
ACCESSION   AE007477 AE005672
VERSION     AE007477.1  GI:14973352
KEYWORDS    .
SOURCE      Streptococcus pneumoniae TIGR4.
  ORGANISM  Streptococcus pneumoniae TIGR4
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 12841)
  AUTHORS   Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
            Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
            Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
            Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D.,
            Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L.,
            McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T.,
            Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O.,
            Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and
            Fraser,C.M.
  TITLE     Complete genome sequence of a virulent isolate of Streptococcus
            pneumoniae
  JOURNAL   Science. 293 (5529), 498-506 (2001)
  MEDLINE   21357209
   PUBMED   11463916
REFERENCE   2  (bases 1 to 12841)
  AUTHORS   Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
            Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
            Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
            Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E.,
            Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A.,
            Feldblyum,T.V., Angiuoli,S., Gesuwan,P., Hickey,E.K., Holt,I.E.,
            Loftus,B.J., Ujwal,M.L., Yang,F., Smith,H.O., Venter,J.C.,
            Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .12841
                     /organism="Streptococcus pneumoniae TIGR4"
                     /strain="TIGR4"
                     /db_xref="taxon:170187"
     gene            complement(121. .525)
                     /gene="SP1851"
     CDS             complement(121. .525)
                     /gene="SP1851"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK75923.1"
                     /db_xref="GI:14973353"
                     /translation="MKDFHFDAISAFENYEIEKMRDGHVVVTTKVVNSSLNYYGNAHG
                     GYLFTLCDQISGLVVISLGLDGVTLQSSINYLKAGKLDDVLTIKGECVHQGRTTCVMD
                     VDITNQEGRNVCKATFTMFVTGQRSEERRVRI"
     gene            complement(596. .2077)
                     /gene="SP1852"
     CDS             complement(596. .2077)
                     /gene="SP1852"
                     /note="identified by match to PFAM  protein family HMM
                     PF01087"
                     /codon_start=1
                     /transl_table=11
                     /product="galactose-1-phosphate uridylyltransferase"
                     /protein_id="AAK75924.1"
                     /db_xref="GI:14973354"
                     /translation="MTLVNKFVTHVISESSFEEMDRIYLTNRVLARVGEGVLEVETNL
                     DKLIDLKDQLVEEAVRLETIEDSQTAREILGAELMDLVTPCPSQVNRDFWATYAHSPE
                     QAIEDFYQLSQKNDYIKLKAIARNIAYRVPSDYGELEITINLSKPEKDPKEIVAAKLV
                     QASNYPQCQLCLENEGYHGRVNHPARSNHRIIRFEMVGQEWGFQYSPYAYFNEHCIFL
                     DGQHRPMAISRQSFERLLAIVDQFPGYFAGSNADLPIVGGSILTHDHYQGGRHVFPME
                     LAPLQKAFRFAGFEQVKAGIVKWPMSVLRLTSDSKEDLINLADKILQEWRQYSDPAVQ
                     ILAETDRTPHHTITPIARKRDGQFELDLVLRDNQTSAEYPDGIYHPHKDVQHIKKENI
                     GLIEVMGLAILPPRLKEEVEQVASYLVGEAVTVADYHQEWADQLKSQHPDLTDKEKAL
                     AIVKDSVGAIFARVLEDAGVYKQTEQGQTAFMRFVEQVGILLD"
     gene            complement(2096. .3274)
                     /gene="SP1853"
     CDS             complement(2096. .3274)
                     /gene="SP1853"
                     /note="identified by match to TIGR  protein family HMM
                     TIGR00131"
                     /codon_start=1
                     /transl_table=11
                     /product="galactokinase"
                     /protein_id="AAK75925.1"
```

```
                /db_xref="GI:14973355"
                /translation="MAQHLTTEALRKDFLAVFGQEADQTFFSPGRINLIGEHTDYNGG
                HVFPAAISLGTYGAARKRDDQVLRFYSANFEDKGIIEVPLADLKFEKEHNWTNYPKGV
                LHFLQEAGHVIDKGFDFYVYGNIPNGAGLSSSASLELLTGVVAEHLFDLKLERLDLVK
                IGKQTENNFIGVNSGIMDQFAIGMGADQRAIYLDTNTLEYDLVPLDLKDNVVVIMNTN
                KRRELADSKYNERRAECEKAVEELQVSLDIQTLGELDEWAVDQYSYLIKDENRLKRAR
                HAVLENQRTLKAQVALQAGDLETFGRLMNASHVSLEHDYEVTGLELDTLVHTAWAQEG
                VLGARMTGAGFGGCAIALVQKDTVEAFKEAVGKHYEEVVGYAPSFYIAEVAGGTRVLD
                "
gene            3439. .4446
                /gene="SP1854"
CDS             3439. .4446
                /gene="SP1854"
                /codon_start=1
                /transl_table=11
                /product="galactose operon repressor"
                /protein_id="AAK75926.1"
                /db_xref="GI:14973356"
                /translation="MATLKDIAQLASVSIATVSRVLNRDQSLSVTEETRHRILTVAEE
                LGYTKHLKTGESHKPKQKIAIIQWVSEQGELDDLYYYQIRLGIEKRAQELDYDILRYF
                NDHPFTLSEEVIGILCIGKFSRAQISAFEEYQKPLVFLDSDTLSLGHTCIITDFYTAM
                KQVVDYFLSQGMDRIGILTGLEETTDQEEIIQDKRLENFKNYSQARGIYHDELVFQGR
                FTAQSGYDLMKEAIQSLGDQLPPAFFAASDSLAIGALRALQEAGISLPDRVSLISFND
                TSLTKQVYPPLSSITVYTEEMGRAGMDILNKEVLHGRKIPSLTMLGTRLTLRESTLNQ
                E"
gene            complement(4492. .5529)
                /gene="SP1855"
CDS             complement(4492. .5529)
                /gene="SP1855"
                /codon_start=1
                /transl_table=11
                /product="alcohol dehydrogenase, zinc-containing"
                /protein_id="AAK75927.1"
                /db_xref="GI:14973357"
                /translation="MKSAVYTKAGQVGLASIERPQIIEADDVIIRVVRACVCGSDLWR
                YRNPETKAGHKNSGHEAIGIVEEAGEAITTVKPGDFVIVPFTHGCGECDACLAGFDGS
                CDNHIGNNLGGDFQAEYIRFHYANWALVKIPGQPSDYTEGMLKSLLTLADVMPTGYHA
                ARVANVQKGDKVVVIGDGAVGQCAVIAAKMRGASQIILMSRHEDRQKMAMESGATAVV
                AERGQEGITKVREILGGGADAALECVGTEAAIEQALGVLHNGGRMGFVGVPHYNNRAL
                GSTFMQNISVAGGAASATTYDKQFLLKAVLDGDINPGRVFTSSYKLEDIDQAYKDMDE
                RKTIKSMIVIE"
gene            complement(5530. .5883)
                /gene="SP1856"
CDS             complement(5530. .5883)
                /gene="SP1856"
                /codon_start=1
                /transl_table=11
                /product="transcriptional regulator, MerR family"
                /protein_id="AAK75928.1"
                /db_xref="GI:14973358"
                /translation="MNIKSASDLLGISADTIRYYERVGLVPPITRTATGIRDFQDQDI
                EALEFIKCFRSAGVSVDSLVDYMSLYQKGDETREERLGILEEEKQKLEERLSQLQTAL
                NRLNLKIKLYKEGKF"
gene            complement(6127. .7026)
                /gene="SP1857"
CDS             complement(6127. .7026)
                /gene="SP1857"
                /note="identified by match to PFAM  protein family HMM
                PF01545"
                /codon_start=1
                /transl_table=11
                /product="cation efflux system protein"
                /protein_id="AAK75929.1"
                /db_xref="GI:14973359"
                /translation="MRNMKAKYAVWVAFFLNLTYAIVEFIAGGVFGSSAVLADSVHDL
                GDAIAIGISAFLETISNREEDNQYTLGYKRFSLLGALVTAVILVTGSVLVILENVTKI
                LHPQPVNDEGILWLGIIAITINLLASLVVGKGKTKNESILSLHFLEDTLGWVAVILMA
                IVLRFTDWYILDPLLSLVISFFILSKALPRFWSTLKIFLDAVPEGLDIKQVKSGLERL
                DNVASLNQLNLWTMDALEKNAIVHVCLKEMEHMETCKESIRIFLKDCGFQNITIEIDA
                DLETHQTHKRKVCDLERSYEHQH"
gene            7150. .7686
                /gene="SP1858"
CDS             7150. .7686
                /gene="SP1858"
                /codon_start=1
                /transl_table=11
                /product="transcriptional regulator, TetR family"
                /protein_id="AAK75930.1"
                /db_xref="GI:14973360"
                /translation="MTNIDRRISKTKKAIYQAFIQLLNAKGYEATTVQDIIDLADVGR
                STFYCHYESKELLLDQLCRYLFHHLFEREQAISTEDYLAHLFLHFHKNQDHITSLLFS
                KNDYFLRQLHKELEHHVYSVLADNLKEAHPNLPTSYLQHLVMSNFIETLTWWLKKGQD
                FTDQEVVQFYLDLLIPKN"
gene            complement(7853. .8662)
                /gene="SP1859"
CDS             complement(7853. .8662)
                /gene="SP1859"
                /note="similar to GB:L42023 PID:1161423 PID:1221194
                PID:1205327 SP:Q57425; identified by sequence similarity;
                putative"
                /codon_start=1
                /transl_table=11
                /product="transporter, putative"
                /protein_id="AAK75931.1"
                /db_xref="GI:14973361"
                /translation="MMHTYLQKKIENIKTTLGEMSGGYRRMVAAMADLGFSGTMKAIW
                DDLFAHRSFAQWIYLLVLGSFPLWLELVYEHRIVDWIGMICSLTGIICVIFVSEGRAS
                NYLFGLINSVIYLILALQKGFYGEVLTTLYFTVMQPIGLLVWIYQAQFKKEKQEFVAR
                KLDGKGWTKYLSISVLWWLAFGFIYQSIGANRPYRDSITDATNGVGQILMTAVYREQW
                IFWAATNVFSIYLWWGESLQIQGKYLIYLINSLVGWYQWSKAAKQNTDLLN"
gene            complement(9004. .10524)
                /gene="SP1860"
CDS             complement(9004. .10524)
                /gene="SP1860"
                /codon_start=1
                /transl_table=11
                /product="choline transporter"
                /protein_id="AAK75932.1"
                /db_xref="GI:14973362"
                /translation="MTNLIATFQDRFSDWLTALSQHLQLSLLTLLLAILLAIPLAVFL
                RYHEKLADWVLQIAGIFQTIPSLALLGLFIPLMGIGTLPALTALVIYAIFPILQNTIT
                GLKGIDPNLQEAGIAFGMTRWERLKKFEIPLAMPVIMSGIRTAAVLIIGTATLAALIG
                AGGLGSFILLGIDRNNASLILIGALSSAVLAIAFNFLLKVMEKAKLRTIFSGFALVAL
                LLGLSYSPALLVQKEKENLVIAGKIGPEPEILANMYKLLIEENTSMTATVKPNFGKTS
                FLYEALKKGDIDIYPEFTGTVTESLLQPSPKVSHEPEQVYQVARDGIAKQDHLAYLKP

alignment_scores:
            Quality:   78.00              Length:    127
              Ratio:  1.200                Gaps:      6
 Percent Similarity:  51.181    Percent Identity: 28.346

alignment_block:
 US-10-048-197-2 x AE007477/rev   ..

 Align seg 1/1  to reverse of: AE007477  from: 1  to: 12841

      10 IleSerAlaLeuLeuThrAla...LeuLeuValThrGlyCysValSerTh 25
         ::::::||||||:::||||||   :::|||||||||||||   |||   ::
    6795 CTAGGAGCCTTGGTAACAGCTGTGATTCTCGTAACGGGCTCTGTTCTAGT 6746

      25 r......GlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
         :      ||||||:::          :::   |||   :::   ::::
    6745 CATTTTGGAAAATGTCACGAAGATTTTGCATCCGCAACCAGTCAATGATG 6696

      40 lnThrIleIle..................................... 43
         ::   |||:::
    6695 AGGGGATTCTCTGGTTAGGAATTATTGCGATTACTATCAATCTGTTAGCG 6646

      44 ..............LysGlyLysThrAsnLysGlnGluIleSerSerAr 55
                       ||||||||||||:::::::::::::|||   |||
    6645 AGTCTGGTGGTTGGTAAGGGAAAGACAAAGAATGAGTCTATTCTGAGTCT 6596

      55 gPheGlySerAlaAspSerIleSerPhe..............MetIleV 67
                 |||:::::::::::::                 |||   :
    6595 GCATTTTCTGGAAGATACGCTAGGGTGGGTAGCTGTTATCCTGATGGCGA 6546
```

```
     67 alValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 83
        ::|||:::::|||            ||||||   |||
   6545 TTGTTCTTCGATTTACGGACTGGTATATCCTAGATCCT..... 6508

     84 GluIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPr 100
        :::||||||||||:::|||||||||:::   ::::::         ||
   6507 ...CTTTTGTCCCTTGTCATTTCTTTCTTTATTCTTTCAAAAGCCCTTCC 6461

    100 oLysAsnLeuSerPheTyrLeuThrAlaLys 110
        |:::         |||:::   |||   |||
   6460 ACGT.........TTTTGGTCTACACTCAAG 6439
```

seq_name: gb_pr:AC012068

seq_documentation_block:

LOCUS AC012068 166516 bp DNA linear PRI 30-SEP-2000
DEFINITION Homo sapiens BAC clone RP11-196J6 from Y, complete sequence.
ACCESSION AC012068
VERSION AC012068.5 GI:10047984
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166516)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 166516)
AUTHORS Andrews,S., Stoneking,T. and Gibson,A.
TITLE The sequence of Homo sapiens BAC clone RP11-196J6
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 166516)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 166516)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 166516)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 166516)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Sep 10, 2000 this sequence version replaced gi:9454626.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0196J06
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-115H13, 200 bp overlap; the clone sequenced to the right is RP11-540C18, 200 bp overlap. Actual start of this clone is at base position 195 of RP11-196J6; actual end is at base position 30203 of RP11-540C18.

FEATURES Location/Qualifiers

source	1. .166516	
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="Y"
		/map="Y"
		/clone="RP11-196J6"
		/clone_lib="RPCI-11"
misc_feature	1. .328	
		/note="similar to EST AW964885 (NID:g8154616)"
misc_feature	76. .336	
		/note="similar to EST AA325338 (NID:g1977613)"
misc_feature	79. .468	
		/note="similar to EST AI471609 (NID:g4333699) ti66a01.x1"
misc_feature	311. .746	
		/note="similar to EST AI810903 (NID:g5397390) tu09a10.x1"
misc_feature	384. .727	
		/note="similar to EST AL134262 (NID:g6602449)"
misc_feature	489. .746	
		/note="similar to EST AW367212 (NID:g6871862)"
misc_feature	976. .1038	
		/note="similar to EST AW957794 (NID:g8147477)"
repeat_region	1061. .1108	
		/rpt_family="L1"
misc_feature	5114. .5681	
		/note="similar to EST AW957794 (NID:g8147477)"
misc_feature	5121. .5650	
		/note="similar to EST AA481204 (NID:g2210756) aa34d03.r1"
misc_feature	5274. .5698	
		/note="similar to EST AA307574 (NID:g1959903)"
misc_feature	5463. .5801	
		/note="similar to EST AI910976 (NID:g5630712) wd22a06.x1"
misc_feature	5861. .6086	
		/note="similar to EST BE169475 (NID:g8632196)"
repeat_region	6083. .6355	
		/rpt_family="ERV1"
repeat_region	6350. .6868	
		/rpt_family="ERV1"
repeat_region	7040. .8710	
		/rpt_family="ERV1"
repeat_region	8712. .8981	
		/rpt_family="ERVK"
misc_feature	8802. .9134	
		/note="similar to EST AI766349 (NID:g5232858) wh60e09.x1"

```
    misc_feature    8802. .9250
                    /note="similar to  EST AI083632 (NID:g3422055) qf26f02.x1"
    misc_feature    8802. .9284
                    /note="similar to  EST AA954654 (NID:g3118349) om95b10.s1"
    misc_feature    8803. .9285
                    /note="similar to  EST AI199588 (NID:g3752194) qi57d03.x1"
    misc_feature    8869. .9241
                    /note="similar to  EST AA909815 (NID:g3050614) ok81g08.s1"
    misc_feature    8932. .9352
                    /note="similar to  EST AW902220 (NID:g8066425)"
    misc_feature    8968. .9583
                    /note="similar to  EST AI174899 (NID:g6361302)"
    repeat_region   9090. .9725
                    /rpt_family="ERVK"
    repeat_region   13346. .15412
                    /rpt_family="ERVK"
    repeat_region   15448. .15655
                    /rpt_family="ERVK"
    repeat_region   15656. .15925
                    /rpt_family="ERV1"
    repeat_region   15952. .16126
                    /rpt_family="ERV1"
    repeat_region   16144. .16222
                    /rpt_family="ERV1"
    repeat_region   16380. .16661
                    /rpt_family="Alu"
    repeat_region   17707. .17995
                    /rpt_family="Alu"
    repeat_region   18008. .18061
                    /rpt_family="Alu"
    repeat_region   18197. .18548
                    /rpt_family="MaLR"
    repeat_region   18563. .18790
                    /rpt_family="L1"
    repeat_region   20146. .20209
                    /rpt_family="L1"
    repeat_region   20428. .20524
                    /rpt_family="L1"
    repeat_region   20527. .20568
                    /rpt_family="MIR"
    repeat_region   20599. .20902
                    /rpt_family="Alu"
    repeat_region   21945. .22219
                    /rpt_family="Alu"
    repeat_region   22224. .22332
                    /rpt_family="AcHobo"
    repeat_region   22656. .22703
                    /rpt_family="AcHobo"
    repeat_region   24332. .24394
                    /rpt_family="MER1_type"
    repeat_region   24568. .24682
                    /rpt_family="Alu"
    misc_feature    24909. .25046
                    /note="similar to  EST AI760835 (NID:g5176502) wh96f11.x1"
    repeat_region   26262. .26545
                    /rpt_family="Alu"
    repeat_region   27732. .27833
                    /rpt_family="MER1_type"
    repeat_region   29260. .29541
                    /rpt_family="Alu"
    repeat_region   29945. .29998
                    /rpt_family="MaLR"
    repeat_region   29999. .30278
                    /rpt_family="Alu"
    repeat_region   30279. .30463
                    /rpt_family="MaLR"
    repeat_region   31151. .31313
                    /rpt_family="MER1_type"
    repeat_region   31473. .31611

alignment_scores:
          Quality:   78.00          Length:    105
            Ratio:  1.279            Gaps:      4
 Percent Similarity:  58.095   Percent Identity: 24.762

alignment_block:
 US-10-048-197-2 x AC012068/rev   ..

 Align seg 1/1  to reverse of: AC012068  from: 1  to: 166516

     11 SerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGlyAs 27
        |||   :::|||      |||:::::::::::::|||   |||:::   ::
  47750 AGCAAATTCCTTTTTTTTTTGGTGATCGCGTCTTGC...TCTGTCACCCA 47704

     27 nValAlaMetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleL 44
        ::::::::::::::   ||||||   ::::::::::::   :::   :::
  47703 GGCTGGAGTGCAGTGGCAAAATCTTGACTCGTTGCAAGGTTCAAGCCTCC 47654

     44 ysGlyLysThrAsnLysGlnGluIleSerSerArgPheGlySerAlaAsp 60
          |||:::::::|||           :::|||:::   ::::::
  47653 CAGGTTCAAGCAATCCTCCTGCCTCAGCCTCCCAAGTAGCTGCGATTACA 47604

     61 Ser.....................IleSerPheMetIleValValIleLy 70
        :::                     |||:::   ::::::|||   |||
  47603 GGCACCTGCCACCACAACCAGCTAATTACTGTATTTTTAGTAGAAATATG 47554

     70 sPheGlyHisThr.AlaIleLeuAlaProAsnArgTrpGlnGluIleLeu 86
         |||   |||::: |||   |||:::   |||  |||   :::::::::
  47553 CTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGGCCTCAAGTGATC 47504

     87 SerLeuIleIleSerPheLeuTrpValLysProTyrArgProLysAsnLe 103
          |||                       |||:::   ||||||   ||
  47503 CACCTG.....................CCCTGGCCTCCCAAAGTGTT 47478

    103 uSerPheTyrLeu 107
        |:::||||||:::
  47477 GGGATTCTACGTG 47465

seq_name: gb_htg:AL356370

seq_documentation_block:
LOCUS       AL356370              87923 bp    DNA     linear   HTG 13-JUN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP5-1008I19, *** SEQUENCING IN
            PROGRESS ***, 22 unordered pieces.
ACCESSION   AL356370
VERSION     AL356370.2  GI:9797509
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 87923)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 12, 2000 this sequence version replaced gi:9213857.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ1008I19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 78098 bases at least Q40
            Consensus quality: 81125 bases at least Q30
            Consensus quality: 83191 bases at least Q20
            Insert size: 85823; sum-of-contigs
            Insert size: 128278; 5.2% error; agarose-fp
            Quality coverage: 2.64x in Q20 bases; sum-of-contigs Quality
            coverage: 2.02x in Q20 bases; agarose-fp
```

```
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7745: contig of 7745 bp in length
            *     7746 7845: gap of      100 bp
            *     7846    10386: contig of 2541 bp in length
            *    10387 10486: gap of      100 bp
            *    10487    14930: contig of 4444 bp in length
            *    14931 15030: gap of      100 bp
            *    15031    19573: contig of 4543 bp in length
            *    19574 19673: gap of      100 bp
            *    19674    22034: contig of 2361 bp in length
            *    22035 22134: gap of      100 bp
            *    22135    24496: contig of 2362 bp in length
            *    24497 24596: gap of      100 bp
            *    24597    28696: contig of 4100 bp in length
            *    28697 28796: gap of      100 bp
            *    28797    37035: contig of 8239 bp in length
            *    37036 37135: gap of      100 bp
            *    37136    39184: contig of 2049 bp in length
            *    39185 39284: gap of      100 bp
            *    39285    45630: contig of 6346 bp in length
            *    45631 45730: gap of      100 bp
            *    45731    47910: contig of 2180 bp in length
            *    47911 48010: gap of      100 bp
            *    48011    50732: contig of 2722 bp in length
            *    50733 50832: gap of      100 bp
            *    50833    52856: contig of 2024 bp in length
            *    52857 52956: gap of      100 bp
            *    52957    54989: contig of 2033 bp in length
            *    54990 55089: gap of      100 bp
            *    55090    58682: contig of 3593 bp in length
            *    58683 58782: gap of      100 bp
            *    58783    63988: contig of 5206 bp in length
            *    63989 64088: gap of      100 bp
            *    64089    67293: contig of 3205 bp in length
            *    67294 67393: gap of      100 bp
            *    67394    71607: contig of 4214 bp in length
            *    71608 71707: gap of      100 bp
            *    71708    74637: contig of 2930 bp in length
            *    74638 74737: gap of      100 bp
            *    74738    78160: contig of 3423 bp in length
            *    78161 78260: gap of      100 bp
            *    78261    80922: contig of 2662 bp in length
            *    80923 81022: gap of      100 bp
            *    81023    87923: contig of 6901 bp in length.
FEATURES             Location/Qualifiers
     source          1. .87923
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP5-1008I19"
                     /clone_lib="RPCI-5"
     misc_feature    1. .7745
                     /note="assembly_fragment:00532
                     fragment_chain:1"
     misc_feature    7846. .10386
                     /note="assembly_fragment:00251
                     fragment_chain:1"
     misc_feature    10487. .14930
                     /note="assembly_fragment:00022
                     fragment_chain:1"
     misc_feature    15031. .19573
                     /note="assembly_fragment:00483
                     fragment_chain:1"
     misc_feature    19674. .22034
                     /note="assembly_fragment:00432
                     fragment_chain:1"
     misc_feature    22135. .24496
                     /note="assembly_fragment:00279
                     fragment_chain:1"
     misc_feature    24597. .28696
                     /note="assembly_fragment:00401
                     fragment_chain:2"
     misc_feature    28797. .37035
                     /note="assembly_fragment:00446
                     fragment_chain:2"
     misc_feature    37136. .39184
                     /note="assembly_fragment:00731
                     fragment_chain:3"
     misc_feature    39285. .45630
                     /note="assembly_fragment:00690
                     fragment_chain:3"
     misc_feature    45731. .47910
                     /note="assembly_fragment:00012"
     misc_feature    48011. .50732
                     /note="assembly_fragment:00039"
     misc_feature    50833. .52856
                     /note="assembly_fragment:00087"
     misc_feature    52957. .54989
                     /note="assembly_fragment:00132"
     misc_feature    55090. .58682
                     /note="assembly_fragment:00156"
     misc_feature    58783. .63988
                     /note="assembly_fragment:00206"
     misc_feature    64089. .67293
                     /note="assembly_fragment:00227"
     misc_feature    67394. .71607
                     /note="assembly_fragment:00301"
     misc_feature    71708. .74637
                     /note="assembly_fragment:00472"
     misc_feature    74738. .78160
                     /note="assembly_fragment:00608"
     misc_feature    78261. .80922
                     /note="assembly_fragment:00702"
     misc_feature    81023. .87923
                     /note="assembly_fragment:00634
                     clone_end:T7
                     vector_side:right"
BASE COUNT    24752 a  18508 c  17954 g  24603 t   2106 others
ORIGIN


alignment_scores:
           Quality:   76.50            Length:    104
             Ratio:  1.275              Gaps:      3
 Percent Similarity:  57.692   Percent Identity: 25.962

alignment_block:
 US-10-048-197-2 x AL356370   ..

 Align seg 1/1  to: AL356370  from: 1  to: 87923

     19 ValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGl 35
        |||   |||       |||   |||       |||          ::::::
  55372 GTACCAGGAGGACAAAGCTGTGGCCCGTGGGCCACTTTGACCCACAAATG 55421

     35 nGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln.... 50
         ::::::::::::|||:::::::|||::::::   |||   ||||||
  55422 TGATGTGCTTAGCCAAGCACTCATAAGTGCTTAAACTATAAAACAGTTTT 55471

     51 ..............GluIleSerSerArgPheGlySerAlaAspSerIle 62
                     :::   :::|||:::|||:::|||   |||   :::
  55472 TAGTGCCTTTGGCATCATATGCTTCCCAGTTTGCCAGT...GATCCTCTT 55518

     63 SerPheMetIleValValIleLysPheGlyHis.................. 73
        |||:::   ::::::          |||   |||
  55519 TCTTATTGTCTTATACCAAGCCCCTTCTAGCACATGACTTGCTATGCATC 55568

     74 .......ThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerL 88
```

```
              :::|||   ::::::   :::   |||::::::   |||:::|
   55569 AAGCCTCAGTGCAGGCTTCACTTCCCATCTCTGGAGTAGTTGACTGGAAC 55618

      88 euIleIleSerPheLeuTrpValLysProTyrArgProLysAsnLeuSer 104
         ||   :::|||      |||:::   |||:::   |||::::::|||:::
   55619 TCTACCTGAGTGGGGATTGGATCATGCCTTTTCCTCCATCATCCCTCGAT 55668

     105 PheTyrLeuThr 108
             :::|||:::
   55669 GATTTTCTGAGC 55680

seq_name: gb_htg:AL138848

seq_documentation_block:
LOCUS       AL138848              102351 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP5-881J12 map p32.2-33, ***
            SEQUENCING IN PROGRESS ***, 10 unordered pieces.
ACCESSION   AL138848
VERSION     AL138848.4  GI:9926433
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 102351)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 25, 2000 this sequence version replaced gi:9212048.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ881J12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 95155 bases at least Q40
            Consensus quality: 97670 bases at least Q30
            Consensus quality: 99255 bases at least Q20
            Insert size: 101451; sum-of-contigs
            Insert size: 120777; 2.6% error; agarose-fp
            Quality coverage: 3.57x in Q20 bases; sum-of-contigs Quality
            coverage: 3.04x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10845: contig of 10845 bp in length
            *    10846 10945: gap of      100 bp
            *    10946    22122: contig of 11177 bp in length
            *    22123 22222: gap of      100 bp
            *    22223    35055: contig of 12833 bp in length
            *    35056 35155: gap of      100 bp
            *    35156    50124: contig of 14969 bp in length
            *    50125 50224: gap of      100 bp
            *    50225    60215: contig of 9991 bp in length
            *    60216 60315: gap of      100 bp
            *    60316    66576: contig of 6261 bp in length
            *    66577 66676: gap of      100 bp
            *    66677    70409: contig of 3733 bp in length
            *    70410 70509: gap of      100 bp
            *    70510    79000: contig of 8491 bp in length
            *    79001 79100: gap of      100 bp
            *    79101    89688: contig of 10588 bp in length
            *    89689 89788: gap of      100 bp
            *    89789   102351: contig of 12563 bp in length.
FEATURES             Location/Qualifiers
     source          1. .102351
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p32.2-33"
                     /clone="RP5-881J12"
                     /clone_lib="RPCI-5"
     misc_feature    1. .10845
                     /note="assembly_fragment:00700
                     clone_end:SP6
                     vector_side:left"
     misc_feature    10946. .22122
                     /note="assembly_fragment:00611
                     fragment_chain:1"
     misc_feature    22223. .35055
                     /note="assembly_fragment:00016
                     fragment_chain:1"
     misc_feature    35156. .50124
                     /note="assembly_fragment:00482
                     fragment_chain:1"
     misc_feature    50225. .60215
                     /note="assembly_fragment:00206
                     fragment_chain:2"
     misc_feature    60316. .66576
                     /note="assembly_fragment:00367
                     fragment_chain:2"
     misc_feature    66677. .70409
                     /note="assembly_fragment:00244"
     misc_feature    70510. .79000
                     /note="assembly_fragment:00869
                     fragment_chain:3"
     misc_feature    79101. .89688
                     /note="assembly_fragment:00779
                     fragment_chain:3"
     misc_feature    89789. .102351
                     /note="assembly_fragment:00186
                     fragment_chain:3
                     clone_end:T7
                     vector_side:right"
BASE COUNT    29645 a  21820 c  21315 g  28664 t    907 others
ORIGIN


alignment_scores:
            Quality:   76.50            Length:    104
              Ratio:  1.275              Gaps:      3
 Percent Similarity:  57.692   Percent Identity: 25.962

alignment_block:
 US-10-048-197-2 x AL138848   ..

 Align seg 1/1  to: AL138848  from: 1  to: 102351

     19 ValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGl 35
        |||   |||      |||   |||      |||         ::::::
   98031 GTACCAGGAGGACAAAGCTGTGGCCCGTGGGCCACTTTGACCCACAAATG 98080

     35 nGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln.... 50
        ::::::::::::|||::::::|||::::::   |||   ||||||
   98081 TGATGTGCTTAGCCAAGCACTCATAAGTGCTTAAACTATAAAACAGTTTT 98130

     51 .............GluIleSerSerArgPheGlySerAlaAspSerIle 62
                     :::   :::|||:::|||:::|||   |||   :::
   98131 TAGTGCCTTTGGCATCATATGCTTCCCAGTTTGCCAGT...GATCCTCTT 98177

     63 SerPheMetIleValValIleLysPheGlyHis.................. 73
        |||:::   ::::::         |||   |||
   98178 TCTTATTGTCTTATACCAAGCCCCTTCTAGCACATGACTTGCTATGCATC 98227
```

```
     74 .......ThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerL 88
                ::|||   ::::::   :::   |||:::::   |||:::|
  98228 AAGCCTCAGTGCAGGCTTCACTTCCCATCTCTGGAGTAGTTGACTGGAAC 98277

     88 euIleIleSerPheLeuTrpValLysProTyrArgProLysAsnLeuSer 104
        ||   :::|||      |||:::   |||:::   |||::::::|||:::
  98278 TCTACCTGAGTGGGGATTGGATCATGCCTTTTCCTCCATCATCCCTCGAT 98327

    105 PheTyrLeuThr 108
          :::|||:::
  98328 GATTTTCTGAGC 98339

seq_name: gb_pr:AL162430

seq_documentation_block:
LOCUS       AL162430              211791 bp    DNA     linear   PRI 30-NOV-2000
DEFINITION  Human DNA sequence from clone RP11-296A18 on chromosome 1, complete
            sequence.
ACCESSION   AL162430
VERSION     AL162430.15  GI:11544472
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 211791)
  AUTHORS   Wray,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 4, 2000 this sequence version replaced gi:11493251.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-296A18 is from the library RPCI-11.2 constructed at the
            Roswell Park Cancer Institute by the group of Pieter de Jong. For
            further details see http://bacpac.med.buffalo.edu/
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-296A18.
FEATURES             Location/Qualifiers
     source          1. .211791
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-296A18"
                     /clone_lib="RPCI-11.2"
BASE COUNT    61174 a  45199 c  44010 g  61408 t
ORIGIN


alignment_scores:
            Quality:   76.50             Length:    104
              Ratio:  1.275              Gaps:        3
 Percent Similarity:  57.692   Percent Identity: 25.962
```

```
alignment_block:
 US-10-048-197-2 x AL162430   ..

 Align seg 1/1  to: AL162430  from: 1  to: 211791

     19 ValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGl 35
        |||   |||      |||   |||      |||         ::::::
 186669 GTACCAGGAGGACAAAGCTGTGGCCCGTGGGCCACTTTGACCCACAAATG 186718

     35 nGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln.... 50
         ::::::::::::|||::::::|||::::::   |||   ||||||
 186719 TGATGTGCTTAGCCAAGCACTCATAAGTGCTTAAACTATAAAACAGTTTT 186768

     51 .............GluIleSerSerArgPheGlySerAlaAspSerIle 62
                     :::   :::|||:::|||:::|||   |||   :::
 186769 TAGTGCCTTTGGCATCATATGCTTCCCAGTTTGCCAGT...GATCCTCTT 186815

     63 SerPheMetIleValValIleLysPheGlyHis.................. 73
        |||:::   ::::::         |||   |||
 186816 TCTTATTGTCTTATACCAAGCCCCTTCTAGCACATGACTTGCTATGCATC 186865

     74 .......ThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerL 88
                ::|||   ::::::   :::   |||:::::   |||:::|
 186866 AAGCCTCAGTGCAGGCTTCACTTCCCATCTCTGGAGTAGTTGACTGGAAC 186915

     88 euIleIleSerPheLeuTrpValLysProTyrArgProLysAsnLeuSer 104
        ||   :::|||      |||:::   |||:::   |||::::::|||:::
 186916 TCTACCTGAGTGGGGATTGGATCATGCCTTTTCCTCCATCATCCCTCGAT 186965

    105 PheTyrLeuThr 108
          :::|||:::
 186966 GATTTTCTGAGC 186977

seq_name: gb_pr:AL137850

seq_documentation_block:
LOCUS       AL137850              125842 bp    DNA     linear   PRI 27-JUN-2001
DEFINITION  Human DNA sequence from clone RP11-56P10 on chromosome 9q31.3-33.3,
            complete sequence.
ACCESSION   AL137850
VERSION     AL137850.12  GI:14575073
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125842)
  AUTHORS   Corby,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 28, 2001 this sequence version replaced gi:14132809.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
```

```
          chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
          Group.  Further information can be found at
          http://www.sanger.ac.uk/HGP/Chr9
          RP11-56P10 is from the library RPCI-11.1 constructed by the group
          of Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pBACe3.6
          IMPORTANT: This sequence is not the entire insert of clone
          RP11-56P10 It may be shorter because we sequence overlapping
          sections only once, except for a 100 base overlap.
          The true left end of clone RP11-4H20 is at 125743 in this sequence.
          The true right end of clone RP11-534I8 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .125842
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="q31.3-33.3"
                     /clone="RP11-56P10"
                     /clone_lib="RPCI-11.1"
     misc_feature    complement(1. .98)
                     /note="match: GSS: Em:AQ673337"
     repeat_region   102. .159
                     /note="MER58 repeat: matches 1. .58 of consensus"
     misc_feature    170. .489
                     /note="match: GSS: Em:AQ101073"
     repeat_region   1213. .1372
                     /note="MIR repeat: matches 27. .191 of consensus"
     repeat_region   2717. .2951
                     /note="MLT1A2 repeat: matches 339. .566 of consensus"
     repeat_region   3132. .3234
                     /note="MLT1G repeat: matches 68. .166 of consensus"
     repeat_region   3313. .3367
                     /note="MIR repeat: matches 84. .144 of consensus"
     repeat_region   4880. .5117
                     /note="MIR repeat: matches 2. .249 of consensus"
     repeat_region   5505. .5659
                     /note="L2 repeat: matches 2089. .2258 of consensus"
     repeat_region   6076. .6140
                     /note="MIR repeat: matches 186. .247 of consensus"
     repeat_region   6178. .6478
                     /note="AluSx repeat: matches 1. .309 of consensus"
     repeat_region   6575. .6861
                     /note="AluSc repeat: matches 1. .288 of consensus"
     repeat_region   7014. .7080
                     /note="MIR repeat: matches 18. .82 of consensus"
     misc_feature    complement(8005. .8534)
                     /note="match: GSS: Em:B99058"
     repeat_region   8335. .8696
                     /note="MLT1A2 repeat: matches 1. .366 of consensus"
     repeat_region   9093. .9338
                     /note="HAL1 repeat: matches 551. .806 of consensus"
     repeat_region   9570. .9880
                     /note="AluY repeat: matches 1. .310 of consensus"
     repeat_region   9952. .10117
                     /note="MER5A repeat: matches 10. .189 of consensus"
     misc_feature    10433. .10794
                     /note="match: GSS: Em:AQ076307"
     repeat_region   10523. .10712
                     /note="L2 repeat: matches 2519. .2707 of consensus"
     repeat_region   10847. .11025
                     /note="MER5A repeat: matches 1. .179 of consensus"
     repeat_region   11464. .11692
                     /note="L1MC/D repeat: matches 5318. .5560 of consensus"
     repeat_region   12528. .12722
                     /note="MIR repeat: matches 29. .231 of consensus"
     misc_feature    12887. .13479
                     /note="match: GSS: Em:B71595"
     repeat_region   14082. .14207
                     /note="MIR repeat: matches 14. .140 of consensus"
     repeat_region   14669. .14764
                     /note="MIR repeat: matches 3. .101 of consensus"
     repeat_region   14853. .14924
                     /note="18 copies 4 mer ttct 66% conserved"
     repeat_region   15221. .17233
                     /note="L1ME2 repeat: matches 3851. .5933 of consensus"
     repeat_region   17392. .17465
                     /note="L1MA9 repeat: matches 6195. .6271 of consensus"
     repeat_region   17726. .18368
                     /note="L1ME1 repeat: matches 5534. .6159 of consensus"
     repeat_region   18371. .21223
                     /note="L1ME2 repeat: matches 379. .3697 of consensus"
     repeat_region   22015. .22462
                     /note="L2 repeat: matches 2267. .2750 of consensus"
     misc_feature    24923. .25135
                     /note="match: GSS: Em:AZ660619"
     misc_feature    25144. .25424
                     /note="match: GSS: Em:AZ660619"
     repeat_region   25253. .25336
                     /note="MIR repeat: matches 53. .139 of consensus"
     repeat_region   27221. .27373
                     /note="MIR repeat: matches 98. .250 of consensus"
     repeat_region   29278. .29540
                     /note="AluSx repeat: matches 1. .287 of consensus"
     misc_feature    complement(29378. .29787)
                     /note="match: GSS: Em:AQ375706"
     repeat_region   29541. .29572
                     /note="8 copies 4 mer gaga 96% conserved"
     misc_feature    complement(29789. .30179)
                     /note="match: GSS: Em:AQ242025"
     repeat_region   30134. .30253
                     /note="MIR repeat: matches 56. .193 of consensus"
     repeat_region   30414. .30521
                     /note="MIR repeat: matches 4. .114 of consensus"
     repeat_region   30607. .30836
                     /note="MIR repeat: matches 23. .256 of consensus"
     repeat_region   32111. .32223
                     /note="L2 repeat: matches 2639. .2742 of consensus"
     repeat_region   32765. .32841
                     /note="L2 repeat: matches 2423. .2498 of consensus"
     repeat_region   32996. .33184
                     /note="MER5A repeat: matches 1. .189 of consensus"
     repeat_region   33275. .33575
                     /note="MLT1I repeat: matches 89. .386 of consensus"
     repeat_region   33900. .33948
                     /note="L2 repeat: matches 2702. .2750 of consensus"
     repeat_region   34847. .35022
                     /note="MIR repeat: matches 74. .258 of consensus"
     repeat_region   35024. .35271
                     /note="MIR repeat: matches 17. .262 of consensus"
     repeat_region   35487. .35799
                     /note="AluSx repeat: matches 2. .310 of consensus"
     repeat_region   35826. .35964
                     /note="MIR repeat: matches 98. .247 of consensus"
     repeat_region   36356. .36430
                     /note="MIR repeat: matches 47. .130 of consensus"
     repeat_region   36500. .38254
                     /note="L1PA5 repeat: matches 4386. .6143 of consensus"
     repeat_region   38427. .38540
                     /note="L1ME1 repeat: matches 6042. .6157 of consensus"
     repeat_region   38541. .38683
                     /note="FLAM_C repeat: matches 1. .133 of consensus"
     repeat_region   38684. .38699
                     /note="L1ME1 repeat: matches 6027. .6042 of consensus"
     repeat_region   38700. .38994
                     /note="AluJo repeat: matches 4. .298 of consensus"
     repeat_region   38995. .39666
                     /note="L1ME1 repeat: matches 5369. .6027 of consensus"
     repeat_region   39668. .39712
                     /note="L1M repeat: matches 5322. .5365 of consensus"
     repeat_region   39709. .41178
                     /note="L1P5 repeat: matches 4204. .5645 of consensus"
     repeat_region   41179. .42394
                     /note="L1M4 repeat: matches 4121. .5332 of consensus"
     repeat_region   42395. .42697
                     /note="MER7A repeat: matches 1. .320 of consensus"
```

```
     repeat_region   42715. .43076
                     /note="L1M4 repeat: matches 3759. .4118 of consensus"
     repeat_region   43406. .43521
                     /note="MIR repeat: matches 47. .160 of consensus"
     misc_feature    45209. .46259
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    complement(45984. .46337)
                     /note="match: STS: Em:G25199"
     repeat_region   46585. .46624
                     /note="20 copies 2 mer ca 97% conserved"
     repeat_region   51073. .51110
                     /note="19 copies 2 mer cc 78% conserved"
     misc_feature    complement(51087. .51290)
                     /note="match: GSS: Em:AQ720148"
     misc_feature    51316. .51798
                     /note="match: GSS: Em:AQ121643"
     repeat_region   51859. .51908

alignment_scores:
            Quality:   75.50            Length:    117
              Ratio:  1.198              Gaps:      4
 Percent Similarity:  53.846   Percent Identity: 26.496

alignment_block:
 US-10-048-197-2 x AL137850   ..

 Align seg 1/1  to: AL137850  from: 1  to: 125842

       4 HisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValTh 20
         |||:::      |||   ||||||   ::::::   |||:::|||
  122335 CATCAGCCCTTTTTGGAGATCTCTTTAATTATCCTAGCAATTCTC...TA 122381

      20 rGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnT 37
          ||||||   ::::::   :::::::::   :::|||   :::   |||
  122382 AGGATGCTATGATAATCCCCATATCACAGATGAGGAAACTGAGGCTCAGA 122431

      37 hrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSer 53
            :::::::|||      :::||||||   :::::::|||:::::::::
  122432 AAGGTAAAATGACTTGAGAGTCAGGTAAAAGGCAGAGCCAAGATTTGAAC 122481

      54 SerArgPheGlySerAlaAspSerIleSerPheMetIleValValIle.. 69
         |||      :::   |||   |||   :::||||||   |||:::
  122482 CCAAGACCACAGAAACAAGATCCTATATTCTTAATGATTTTTGTTGTGTA 122531

      70 ................................LysPheGlyHisThrA 75
                                         |||:::      :
  122532 ACAGAGAAAAAACTTTTTCTTTCTTTCTTCCTTCCTTTTTTCTTCTCCTT 122581

      75 laIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIleSer 91
         ::::::|||   |||:::           :::   |||:::::::   |||
  122582 CCTTCCTTCCTCCCTCC.............CTCTCCTCTTTCTTTCCTTCC 122619

      92 PheLeuTrpValLysProTyrArgProLysAsnLeuSerPheTyrLeuTh 108
         ||||||                      :::||||||||||:::|||::
  122620 TTCCTT.......................TCTCTTTCTTTCTTTCTTTC 122645

     108 r 108
         :
  122646 T 122646

seq_name: gb_htg:AC011161

seq_documentation_block:
LOCUS       AC011161              200532 bp    DNA     linear   HTG 30-MAR-2000
DEFINITION  Homo sapiens clone RP11-9L6, WORKING DRAFT SEQUENCE, 33 unordered
            pieces.
ACCESSION   AC011161
VERSION     AC011161.3  GI:7341865
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 200532)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-9L6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200532)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 30, 2000 this sequence version replaced gi:6139126.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3019
                Center clone name: 9_L_6
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 175452 bases at least Q40
                Consensus quality: 186874 bases at least Q30
                Consensus quality: 191961 bases at least Q20
                Insert size: 200000; agarose-fp
                Insert size: 197332; sum-of-contigs
                Quality coverage: 3.2 in Q20 bases; agarose-fp
                Quality coverage: 3.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1310: contig of 1310 bp in length
            *     1311 1410: gap of      100 bp
            *     1411     3606: contig of 2196 bp in length
            *     3607 3706: gap of      100 bp
            *     3707     5209: contig of 1503 bp in length
            *     5210 5309: gap of      100 bp
            *     5310     7094: contig of 1785 bp in length
            *     7095 7194: gap of      100 bp
            *     7195     9372: contig of 2178 bp in length
            *     9373 9472: gap of      100 bp
            *     9473    11334: contig of 1862 bp in length
            *    11335 11434: gap of      100 bp
            *    11435    13898: contig of 2464 bp in length
            *    13899 13998: gap of      100 bp
            *    13999    15543: contig of 1545 bp in length
            *    15544 15643: gap of      100 bp
            *    15644    17721: contig of 2078 bp in length
            *    17722 17821: gap of      100 bp
```

```
            *    17822    20076: contig of 2255 bp in length
            *    20077 20176: gap of      100 bp
            *    20177    22360: contig of 2184 bp in length
            *    22361 22460: gap of      100 bp
            *    22461    25056: contig of 2596 bp in length
            *    25057 25156: gap of      100 bp
            *    25157    29276: contig of 4120 bp in length
            *    29277 29376: gap of      100 bp
            *    29377    33914: contig of 4538 bp in length
            *    33915 34014: gap of      100 bp
            *    34015    39784: contig of 5770 bp in length
            *    39785 39884: gap of      100 bp
            *    39885    43727: contig of 3843 bp in length
            *    43728 43827: gap of      100 bp
            *    43828    48356: contig of 4529 bp in length
            *    48357 48456: gap of      100 bp
            *    48457    53189: contig of 4733 bp in length
            *    53190 53289: gap of      100 bp
            *    53290    58926: contig of 5637 bp in length
            *    58927 59026: gap of      100 bp
            *    59027    64871: contig of 5845 bp in length
            *    64872 64971: gap of      100 bp
            *    64972    70550: contig of 5579 bp in length
            *    70551 70650: gap of      100 bp
            *    70651    76382: contig of 5732 bp in length
            *    76383 76482: gap of      100 bp
            *    76483    82846: contig of 6364 bp in length
            *    82847 82946: gap of      100 bp
            *    82947    91987: contig of 9041 bp in length
            *    91988 92087: gap of      100 bp
            *    92088    98742: contig of 6655 bp in length
            *    98743 98842: gap of      100 bp
            *    98843   105574: contig of 6732 bp in length
            *   105575 105674: gap of      100 bp
            *   105675   114985: contig of 9311 bp in length
            *   114986 115085: gap of      100 bp
            *   115086   124472: contig of 9387 bp in length
            *   124473 124572: gap of      100 bp
            *   124573   134778: contig of 10206 bp in length
            *   134779 134878: gap of      100 bp
            *   134879   150202: contig of 15324 bp in length
            *   150203 150302: gap of      100 bp
            *   150303   166267: contig of 15965 bp in length
            *   166268 166367: gap of      100 bp
            *   166368   183386: contig of 17019 bp in length
            *   183387 183486: gap of      100 bp
            *   183487   200532: contig of 17046 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200532
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-9L6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1310
                     /note="assembly_fragment"
     misc_feature    1411. .3606
                     /note="assembly_fragment"
     misc_feature    3707. .5209
                     /note="assembly_fragment"
     misc_feature    5310. .7094
                     /note="assembly_fragment"
     misc_feature    7195. .9372
                     /note="assembly_fragment"
     misc_feature    9473. .11334
                     /note="assembly_fragment"
     misc_feature    11435. .13898
                     /note="assembly_fragment"
     misc_feature    13999. .15543
                     /note="assembly_fragment"
     misc_feature    15644. .17721
                     /note="assembly_fragment"
     misc_feature    17822. .20076
                     /note="assembly_fragment"
     misc_feature    20177. .22360
                     /note="assembly_fragment"
     misc_feature    22461. .25056
                     /note="assembly_fragment"
     misc_feature    25157. .29276
                     /note="assembly_fragment"
     misc_feature    29377. .33914
                     /note="assembly_fragment"
     misc_feature    34015. .39784
                     /note="assembly_fragment"
     misc_feature    39885. .43727
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    43828. .48356
                     /note="assembly_fragment"
     misc_feature    48457. .53189
                     /note="assembly_fragment"
     misc_feature    53290. .58926
                     /note="assembly_fragment"
     misc_feature    59027. .64871
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    64972. .70550
                     /note="assembly_fragment"
     misc_feature    70651. .76382
                     /note="assembly_fragment"
     misc_feature    76483. .82846
                     /note="assembly_fragment"
     misc_feature    82947. .91987
                     /note="assembly_fragment"
     misc_feature    92088. .98742
                     /note="assembly_fragment"
     misc_feature    98843. .105574
                     /note="assembly_fragment"
     misc_feature    105675. .114985
                     /note="assembly_fragment"
     misc_feature    115086. .124472
                     /note="assembly_fragment"
     misc_feature    124573. .134778
                     /note="assembly_fragment"
     misc_feature    134879. .150202
                     /note="assembly_fragment"
     misc_feature    150303. .166267

alignment_scores:
           Quality:   75.50              Length:    117
             Ratio:  1.198                Gaps:      4
 Percent Similarity:  53.846   Percent Identity: 26.496

alignment_block:
 US-10-048-197-2 x AC011161/rev   ..

 Align seg 1/1  to reverse of: AC011161  from: 1  to: 200532

       4 HisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValTh 20
         |||:::      |||   ||||||   ::::::   |||:::|||
  146114 CATCAGCCCTTTTTGGAGATCTCTTTAATTATCCTAGCAATTCTC...TA 146068

      20 rGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnT 37
         ||||||   ::::::   :::::::::   :::|||   :::   |||
  146067 AGGATGCTATGATAATCCCCATATCACAGATGAGGAAACTGAGGCTCAGA 146018

      37 hrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSer 53
           ::::::|||      :::||||||   ::::::|||:::::::::
  146017 AAGGTAAAATGACTTGAGAGTCAGGTAAAAGGCAGAGCCAAGATTTGAAC 145968

      54 SerArgPheGlySerAlaAspSerIleSerPheMetIleValValIle.. 69
         |||      :::   |||   |||   :::|||||   |||:::
  145967 CCAAGACCACAGAAACAAGATCCTATATTCTTAATGATTTTTGTTGTGTA 145918

      70 ..............................LysPheGlyHisThrA 75
```

```
                                                ||:::      :
  145917 ACAGAGAAAAAACTTTTTCTTTCTTTCTTCCTTCCTTTTTTCTTCTCCTT 145868

      75 laIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIleSer 91
         :::::|||   |||:::            :::   |||::::::   |||
  145867 CCTTCCTTCCTCCCTCC...........CTCTCCTCTTTCTTTCCTTCC 145830

      92 PheLeuTrpValLysProTyrArgProLysAsnLeuSerPheTyrLeuTh 108
         ||||||                  :::||||||||||:::||||::
  145829 TTCCTT.....................TCTCTTTCTTTCTTTCTTTC 145804

     108 r 108
         :
  145803 T 145803
```

```
seq_name: gb_ov:AF173857

seq_documentation_block:
LOCUS       AF173857              4920 bp    mRNA    linear   VRT 09-MAY-2000
DEFINITION  Xenopus laevis receptor protein tyrosine phosphatase rho precursor
            (ptprt) mRNA, partial cds.
ACCESSION   AF173857
VERSION     AF173857.1  GI:5738202
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 4920)
  AUTHORS   Johnson,K.G. and Holt,C.E.
  TITLE     Expression of CRYP-alpha, LAR, PTP-delta, and PTP-rho in the
            developing xenopus visual system
  JOURNAL   Mech. Dev. 92 (2), 291-294 (2000)
  MEDLINE   20193505
   PUBMED   10727868
REFERENCE   2  (bases 1 to 4920)
  AUTHORS   Besco,J.A., Johnson,K.G., Frostholm,A., Popesco,M., Burghes,A.H.M.,
            Holt,C.E. and Rotter,A.
  TITLE     Xenopus laevis receptor protein tyrosine phosphatase rho
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 4920)
  AUTHORS   Besco,J.A., Johnson,K.G., Frostholm,A., Popesco,M., Burghes,A.H.M.,
            Holt,C.E. and Rotter,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-1999) Pharmacology, Ohio State University, 333
            West Tenth Avenue, Columbus, OH 43210, USA
FEATURES             Location/Qualifiers
     source          1. .4920
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     gene            1. .4920
                     /gene="ptprt"
     sig_peptide     1336. .1395
                     /gene="ptprt"
     CDS             1336. .>4920
                     /gene="ptprt"
                     /note="similar to Homo sapiens receptor protein tyrosine
                     phosphatase encoded by GenBank Accession Number AF043644;
                     similar to Mus musculus receptor protein tyrosine
                     phosphatase-rho encoded by GenBank Accession Number
                     AF173857"
                     /codon_start=1
                     /product="receptor protein tyrosine phosphatase rho
                     precursor"
                     /protein_id="AAD50295.1"
                     /db_xref="GI:5738203"
                     /translation="MREPICLLLIFLQLFPAAGGQVFAGGCTFEEHYNSCGYSVALGT
                     NGFTWEQVNTWERPPLDPSIPTGSFMVVNSSGFASGQKAHLLLPVLKENDTHCIDFHY
                     YMSGRTSPGTLNVYVKVNGGPQGNPVWNVSGVVTEGWVHAELAISTFWPNFYQIIFEA
                     VSVKDHSGYIAVDEVRVLAHPCRNAPHFLRLQNVEVNVGQNAHFQCIAGGKWSQHDKL
                     WLQQWNGRDTALMVTRVVNHKRFSATVSVGETSQRSISKYRCVLRSSGGSGVSNYAEP
                     VVKEPPTPIAPPELLAVGATYLWIKPNANSIIGDGPIVLKEVEYRTTSGSWAETHVMD
                     GPTYKLWHLDPDVEYEIRVLLTRPGEGGTGPPGPPLTTRTKCADPVHGPQNVELVDIR
                     ARQLTLQWEPFAYSVTRCHSYNLTLSLQFXSLGKERVEAEEVVLTSSHYTLCGLRPHT
                     AVYLRLALCNPEGRKESGEIVVHTEEDVPGPIPLESIQGGPFEEKIYLQWKAPNETNG
                     VISMYEVSYKAVGSLDPSADLSNQKGRIVKLRNETHHLFVGLYPGSTYSFTIRASTSK
                     GFGPPVTARIATKIAAPSLTDSEAEGPLNEPEPTITVWLKPAQSRGAPVSAYQLVVKE
                     ERAQRSGRAADGGGCFPVPVGYRNASNLESSHYFAAELKPLQLAVTQPFTVGDNRTYN
                     GYWNPPLSPVKSYSIYFQALSTANGETKIGCVRLASKGASTQNSNTVEPEKPADSTVQ
                     TAGVIAGMLMFIIILLGAMMTIKRRNVYSYSYYLKLAKKQKETQSGSQREMGLVASDK
                     SSSKLSSTRNEDTFSSCQDVNGFSTSSPCSFSVQSKNLQGKSLLSSCYSDNSHVDMNQ
                     PTLTIQTHPYRTCEPLEMAYPRGQLQPAIRVADLLQHITQMKRGQGYGFKEEYEALPE
                     GQTASWDTAKEDENRNKNRYGNIISYDHSRVRLQQIEEDPHSDYINANYVDGYHRPRY
                     YIATQGPMQETVKDFWRMIWQENSISVVMVTNLVEVGRVKCVRYWPDDTEVYGDIKVS
                     AIETEPLAEYIIRTFTVQKKGSHEIREIRQFHFTSWPDNGVPCYATGLLGFIRQVKFL
                     NPPDAGPITVHCSAGAGRTGCFIAIDIMLDMAENEGVVDIFNCVRELRAQRVNMVQTE
                     EQYVFVHDAILEACLCGNTAIPVCEFRSVYYNISRMDPQTNCSQIKDEF"
     mat_peptide     1396. .>4920
                     /gene="ptprt"
                     /product="receptor protein tyrosine phosphatase rho"
     misc_feature    1399. .1887
                     /gene="ptprt"
                     /note="Region: MAM domain"
     misc_feature    1933. .2115
                     /gene="ptprt"
                     /note="Region: Ig-like domain"
     misc_feature    2125. .2178
                     /gene="ptprt"
                     /note="Region: fibronectin type III repeat"
     misc_feature    2464. .2730
                     /gene="ptprt"
                     /note="Region: fibronectin type III repeat"
     misc_feature    2767. .3042
                     /gene="ptprt"
                     /note="Region: fibronectin type III repeat"
     misc_feature    3088. .3336
                     /gene="ptprt"
                     /note="Region: fibronectin type III repeat"
     misc_feature    3217. .3228
                     /gene="ptprt"
                     /note="proteolytic subtilisin/furin cleavage site"
     misc_feature    3559. .3621
                     /gene="ptprt"
                     /note="Region: transmembrane domain"
     misc_feature    4504. .4809
                     /gene="ptprt"
                     /note="probably catalytic; Region: phosphatase domain"
BASE COUNT     1290 a   1276 c   1190 g   1151 t     13 others
ORIGIN


alignment_scores:
            Quality:   75.00            Length:     90
              Ratio:  1.562              Gaps:      4
 Percent Similarity:  53.333   Percent Identity: 30.000

alignment_block:
 US-10-048-197-2 x AF173857   ..

 Align seg 1/1  to: AF173857  from: 1  to: 4920

      15 ThrAlaLeuLeuValThrGlyCysValSerThrGlyAsnValAlaMetLy 31
         ||||||||||:::||||||      |||
    2017 ACGGCCCTCATGGTGACACGTGTCGTG..................... 2043

      31 sGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrA 48
               |||:::::   ::::::   |||:::      |||:::|||:
    2044 .......AACCACAAACGCTTCTCAGCCACCGTCAGTGTTGGGGAGACAT 2086

      48 snLysGlnGluIleSer.....................SerArgPheGly 57
         ::::::::::||||||                     |||      |||
    2087 CTCAGCGCAGCATCAGCAAATACCGGTGTGTCCTTCGCTCTTCAGGGGGA 2136

      58 SerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHisTh 74
         |||:::   |||   ::::::         |||:::|||         ||
```

```
    2137 TCAGGAGTCTCC...AACTATGCAGAACCAGTNGTGAAAGAGCCTCCCAC 2183

      74 rAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIleS 91
         |   |||      |||           |||:::|||::::::      :
    2184 CCCAATTGCACCCCCT...........GAGCTCCTTGCTGTGGGGGCCA 2221

      91 erPheLeuTrpValLysPro 97
         :::::||||||:::|||||
    2222 CCTACCTGTGGATCAAGCCA 2241

seq_name: gb_pr:AL512598

seq_documentation_block:
LOCUS       AL512598             139491 bp    DNA     linear   PRI 13-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-165A20 on chromosome 10,
            complete sequence.
ACCESSION   AL512598 AC024986
VERSION     AL512598.11  GI:16944073
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 139491)
  AUTHORS   Ashwell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 15, 2001 this sequence version replaced gi:16151438.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-165A20 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-165A20 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-165A20 is at 1 in this sequence.
            The true left end of clone RP11-80K21 is at 137492 in this
            sequence. The true right end of clone RP11-129O7 is at 37219 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1. .139491
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-165A20"
                     /clone_lib="RPCI-11.1"
BASE COUNT    40284 a  28155 c  27520 g  43532 t
ORIGIN
```

```
alignment_scores:
           Quality:   74.50            Length:     69
             Ratio:  1.910              Gaps:      3
 Percent Similarity:  56.522   Percent Identity: 30.435

alignment_block:
 US-10-048-197-2 x AL512598   ..

 Align seg 1/1  to: AL512598  from: 1  to: 139491

      43 IleLysGlyLysThrAsnLysGlnGluIleSerSerArgPheGlySerAl 59
         |||:::   :::|||:::||||||                     |||
   19557 ATATCCCAGCGAACCAAGAAGCAG.....................AGCTG 19585

      59 aAspSerIleSerPheMetIleValValIleLysPheGlyHisThrAlaI 76
          |||   :::      :::   |||:::   ::::::||||||||||:::
   19586 GGATGTGGTA......TTAAAAGTCATACAAAGGTATGGGCATACTGGGC 19629

      76 leLeuAlaPro............AsnArgTrpGlnGluIleLeuSerLeu 88
         :::::::::|||           ::::::|||::::::|||   :::|||
   19630 TCATGGTTCCTTTCAAGGAACTCTCCCAGTGGGAATCTATTGCAGCTCTC 19679

      89 IleIleSerPheLeuTrpValLysProTyrArgProLysAsnLeuSerPh 105
         :::      |||||||||                       |||   ||
   19680 ACATTAAGAGTCCTATGGGTTGCACTTGAAGGAACTGGGTTCCTTGTGTT 19729

     105 eTyrLeu 107
         |||||||
   19730 TTATTTG 19736

seq_name: gb_htg:AP001795

seq_documentation_block:
LOCUS       AP001795             143873 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-711B7 map 11q23, WORKING
            DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION   AP001795
VERSION     AP001795.2  GI:8117467
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-711B7.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143873)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 143,873 genomic DNA of 11q23
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 143873)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-APR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7592910.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: RP11-711B7
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 134262 bases at least Q40
```

```
    Consensus quality: 138796 bases at least Q30
    Consensus quality: 140745 bases at least Q20
    Insert size: 142073; sum-of-contigs
    Quality coverage: 5.49x in Q20 bases; sum-of-contigs
--------------.
NOTE: This is a 'working draft' sequence. It currently consists of
19 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
         1     35204 contig of  35204 bp in length
     35305     47787 contig of  12483 bp in length
     47888     60452 contig of  12565 bp in length
     60553     71377 contig of  10825 bp in length
     71478     83187 contig of  11710 bp in length
     83288     93031 contig of   9744 bp in length
     93132    101197 contig of   8066 bp in length
    101298    108878 contig of   7581 bp in length
    108979    115261 contig of   6283 bp in length
    115362    120164 contig of   4803 bp in length
    120265    125187 contig of   4923 bp in length
    125288    130015 contig of   4728 bp in length
    130116    132802 contig of   2687 bp in length
    132903    134588 contig of   1686 bp in length
    134689    136968 contig of   2280 bp in length
    137069    138355 contig of   1287 bp in length
    138456    140108 contig of   1653 bp in length
    140209    142441 contig of   2233 bp in length
    142542    143873 contig of   1332 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1     35204: contig of 35204 bp in length
*     35205 35304: gap of      100 bp
*     35305     47787: contig of 12483 bp in length
*     47788 47887: gap of      100 bp
*     47888     60452: contig of 12565 bp in length
*     60453 60552: gap of      100 bp
*     60553     71377: contig of 10825 bp in length
*     71378 71477: gap of      100 bp
*     71478     83187: contig of 11710 bp in length
*     83188 83287: gap of      100 bp
*     83288     93031: contig of 9744 bp in length
*     93032 93131: gap of      100 bp
*     93132    101197: contig of 8066 bp in length
*    101198 101297: gap of      100 bp
*    101298    108878: contig of 7581 bp in length
*    108879 108978: gap of      100 bp
*    108979    115261: contig of 6283 bp in length
*    115262 115361: gap of      100 bp
*    115362    120164: contig of 4803 bp in length
*    120165 120264: gap of      100 bp
*    120265    125187: contig of 4923 bp in length
*    125188 125287: gap of      100 bp
*    125288    130015: contig of 4728 bp in length
*    130016 130115: gap of      100 bp
*    130116    132802: contig of 2687 bp in length
*    132803 132902: gap of      100 bp
*    132903    134588: contig of 1686 bp in length
*    134589 134688: gap of      100 bp
*    134689    136968: contig of 2280 bp in length
*    136969 137068: gap of      100 bp
*    137069    138355: contig of 1287 bp in length
*    138356 138455: gap of      100 bp
*    138456    140108: contig of 1653 bp in length
*    140109 140208: gap of      100 bp
*    140209    142441: contig of 2233 bp in length
*    142442 142541: gap of      100 bp
*    142542    143873: contig of 1332 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143873
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q23"
                     /clone="RP11-711B7"
     misc_feature    1. .35204
                     /note="assembly_fragment"
     misc_feature    35305. .47787
                     /note="assembly_fragment"
     misc_feature    47888. .60452
                     /note="assembly_fragment"
     misc_feature    60553. .71377
                     /note="assembly_fragment clone_end:SP6 vector_side:right"
     misc_feature    71478. .83187
                     /note="assembly_fragment"
     misc_feature    83288. .93031
                     /note="assembly_fragment"
     misc_feature    93132. .101197
                     /note="assembly_fragment"
     misc_feature    101298. .108878
                     /note="assembly_fragment"
     misc_feature    108979. .115261
                     /note="assembly_fragment"
     misc_feature    115362. .120164
                     /note="assembly_fragment"
     misc_feature    120265. .125187
                     /note="assembly_fragment"
     misc_feature    125288. .130015
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    130116. .132802
                     /note="assembly_fragment"
     misc_feature    132903. .134588
                     /note="assembly_fragment"
     misc_feature    134689. .136968
                     /note="assembly_fragment"
     misc_feature    137069. .138355
                     /note="assembly_fragment"
     misc_feature    138456. .140108
                     /note="assembly_fragment"
     misc_feature    140209. .142441
                     /note="assembly_fragment"
     misc_feature    142542. .143873
                     /note="assembly_fragment"
BASE COUNT    47187 a  26487 c  25099 g  43300 t   1800 others
ORIGIN


alignment_scores:
            Quality:   74.50             Length:     85
              Ratio:  1.490               Gaps:      3
 Percent Similarity:  58.824   Percent Identity: 25.882

alignment_block:
 US-10-048-197-2 x AP001795/rev   ..

 Align seg 1/1  to reverse of: AP001795  from: 1  to: 143873

      22 CysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIl 38
         |||::::::||||||   |||:::   |||:::   :::|||::::::||
   46709 TGTGCTGCCACAGGGTTTGTGGTGTCAAAGCAGGCATCTCAGTCCAGCAT 46660

      38 eGlu.GlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSer 54
         ||||       ||||||          ::::::      :::
   46659 TGAGGCCTTTCATCATCACAGTCTTTCTATCTGAAGTTTATCTT..... 46616

      55 ArgPheGlySerAlaAspSerIleSerPheMetIleValValIleLysPh 71
                           :::|||||||:::|||:::      :::||
```

```
   46615 ....................CTGTCATTCCTCATCATACCCACAAGATT 46587

      71 eGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerL 88
         ||||||||||:::::::::      :::           :::   :::
   46586 TGGCCATACCTGTGTCATT......ACCTTTGTTGTTGTTGTACATACAC 46543

      88 euIleIleSerPheLeuTrpValLysProTyrArgProLysAsnLeuSer 104
           ::::::   ::::::|||:::   ||||||      :::::::|||:::
   46542 ATGTTATGATTTACATTTGGGCTTTTCCTTATTCCTTTTCCACGTTAGCA 46493

      105 Phe 105
          |||
   46492 TTT 46490
```

seq_name: gb_pl:AP003215

seq_documentation_block:

LOCUS AP003215 154137 bp DNA linear PLN 04-AUG-2001
DEFINITION Oryza sativa genomic DNA, chromosome 1, BAC clone:OSJNBa0089K24.
ACCESSION AP003215
VERSION AP003215.3 GI:15128217
KEYWORDS .
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:OSJNBa0089K24.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 154137)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, BAC clone:OSJNBa0089K24
JOURNAL Published Only in Database (2001) In press
REFERENCE 2 (bases 1 to 154137)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-2001) Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT On Aug 9, 2001 this sequence version replaced gi:13442957.
Genes were predicted from the integrated results of the following: GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against NCBI NonRedundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP. Protein homologies of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding DDBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein.
The orientation of the sequence is from -21M13 to M13Rev of the BAC clone. This sequence of OSJNBa0089K24 clone has an overlap with P0013F10 clone (DDBJ: AP002523) at the position 151,639 to 154,137 of 3' end. The sequence of this clone ends at the position 2,499 of P0013F10. Detailed information on overlap and assembly quality together with annotation of this entry is available at http://rgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES Location/Qualifiers
source 1. .154137
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="OSJNBa0089K24"
gene join(4692. .4907,4987. .5166,5273. .5414,6051. .6472)
/gene="OSJNBa0089K24.1"
CDS join(4692. .4907,4987. .5166,5273. .5414,6051. .6472)
/gene="OSJNBa0089K24.1"
/codon_start=1
/product="putative peroxidase"
/protein_id="BAB62546.1"
/db_xref="GI:15128218"
/translation="MAISEMSALFFLFSALLRSSLVHSQGLQIGFYDNNCPDAEDIVR
STVEKYYNNDATIAPGLLRLHFHDCFVQGCDASVLISGASSERTAPQNFGIRGFEVID
DAKSQLEAVCSGVVSCADILALAARDAVDLTGGPSWSVPLGRRDGRISSASDAKALPS
PADPVSVQRQKFAAQGLTDRAHTIGQTDCIFFRYRLYNFTATGNADPTISPSALPQLR
ALCPPAGDGSRRVALDLGSPGAFDVSFFKNVRDGGAVLESDQRLWGDAATQAAVQSFA
GNVRGLFGLRFSYEFPKAMVRMSSIAVKTGSQGEIRRKCSKFN"
gene 8547. .9122
/gene="OSJNBa0089K24.2"
CDS 8547. .9122
/gene="OSJNBa0089K24.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB62547.1"
/db_xref="GI:15128219"
/translation="MARRPPPQRHHHLVLLALAAFSCLASGTATPTPTTSSQSQQPQ
PAGTLLQPALAADLAARTCWYTVQIKTSCASPWRTSDAVSLAFGDAYRNEVYAARLAG
SSSSPQPSSSAAFERCATDTFRVGGPCGYGVCYLYLRRSGRDGWTPQWVRVYEPTSDT
PSTFYYGDPLPNAVWYGFNRCPRLAASAAAQ"
gene complement(join(13385. .13475,13662. .13754,13888. .13995, 14101. .14158,14320. .14554))
/gene="OSJNBa0089K24.3"
CDS complement(join(13385. .13475,13662. .13754,13888. .13995, 14101. .14158,14320. .14554))
/gene="OSJNBa0089K24.3"
/note="hypothetical protein
similar to Arabidopsis thaliana chromosome 1, F6D8.19"
/codon_start=1
/protein_id="BAB62548.1"
/db_xref="GI:15128220"
/translation="MATTSACCYRPLSAPPLHRRRGPPPQQQQLRTAVGGPRPRPSLQ
VAAAAATGGGQSVKAATDAEFFQTSDTRPIMLFDGVCNLCNGGVRFVQEHDPNRSIRY
IPLQSESGRKLLQRSGRAPDDISSVVLVEKDRSYIKSEAVLRIMGYLNLPFPQLATFL
NFVPLFLRDFAYDNVADNRYAVFGRSESESCEIL"
gene complement(join(15491. .15586,15712. .15776,15919. .15981, 16599. .16647,16744. .16929))
/gene="OSJNBa0089K24.4"
CDS complement(join(15491. .15586,15712. .15776,15919. .15981, 16599. .16647,16744. .16929))
/gene="OSJNBa0089K24.4"
/note="contains ESTs AU100786(C50379),C26898(C50379)
similar to Arabidopsis thaliana chromosome 1, F28N24.7 unknown protein"
/codon_start=1
/protein_id="BAB62549.1"
/db_xref="GI:15128221"
/translation="MAVEEITEGVRNLAVEGEPAAAAAAAGGGGEGAQRRAAGSSSNR
IQVSNTKKPLFFYVNLAKRYMQQHGDVELSALGMAIATVVTVAEILKNNGFAVEKKIR
TSTVEINDESRVRPLQKAKIEIVLEKSEKFDELMAAAAEEREAAEAEEQA"
gene complement(join(21465. .21650,21730. .21828,21923. .21995, 22080. .22190,22287. .22460,22570. .23666,23741. .25162))
/gene="OSJNBa0089K24.5"
CDS complement(join(21465. .21650,21730. .21828,21923. .21995, 22080. .22190,22287. .22460,22570. .23666,23741. .25162))
/gene="OSJNBa0089K24.5"
/note="hypothetical protein
similar to Oryza sativa chromosome 1, P0489A01.2"
/codon_start=1
/protein_id="BAB62550.1"
/db_xref="GI:15128222"
/translation="MADRDEEQILYDTIAEGSSQYWNEEEGNEDPNQYLNEEGNVERD
AEGNQQGHVERDVEGNQEEEASGSQPSVGQKRARGQRGAAKKLEGRHIITEVEEDGRP
SAPAEAAKNYVRHSGWVVRDNVPVSTVYWRRTRARGDHESFVPDSEKEMLWTTMLETF
TLPAGTEDKVKRWTLKKMAEQFQSFKGDLYLKYILKGQTPNFDTFPKLRDHWDEFVAY
KTGEQGKAMMERNKENAAKKKYHHHLGSGGYSVAMPKWEQIEASLIERGIEPATANWP
ERSKFWYYAHGGTLNPADGSLVFGYQIQEAARRLTDVVEASSQGTFRPDRDRDELTLA
LQTPEHPGRTRGKGVIPWKIGFKEDIHTYRSRMRSKRDTEAKIADLEFRVSSYELNMQ
EEVARKVDERMAAHRSHDPQPTIPPVMVSPSGNRSSCASTGQVGSQSMDAMQTQDEST
```
```

```
                CPVDDITQRTPCELHIPFKNLSIKVASGMAIPTDPSGTYHCRPIPAGYSKVEVELVEG
                AYEDLELDYPGGDGETHLRDTCHAIILWRRRYIILPGRQAASRAPSPPAPPSPPQDPA
                PSPPHAPPAPSPPQSPASTPQDPAPTPPRAPTPTPPQAPLPAPSKSRAPPAPPLAHTR
                ATKKAKVDAAKNKDPGYDCTQEELDAYVASEVKRQFKPRSPEKKIPIDPSVRNFFRGM
                SASVKEAIKLSDYERTLKKASSGKSKPVPQLGEQPNQEIEPLVTGKEMTIEQFITDTG
                LTTDQLLGVAPIEKAEVKYMYELGKPLVKPELLQSLPTQMYKFHQLYMEMSATGREMI
                GARIRDTDFLQGDDILWINFRGIYELYQLDALDVSIMSCWILMEIQRARRRRVFDTGF
                IDPRKVNVAMLDQYPQETEDNLVHLLKAQHYKTFILLPYNTEFHWVLLLIDLEACTVN
                VYDSMDKKESTFDKVFELIDRAWYRFRHLVRGKWRERLRRKFKFPCAKQKQGTNLCGY
                YVCEYCHCLADQIITTRELDFIRMRDNLTTHKEFIAAVQEQLMGFINEEILDPKGEFY
                YDGNTIHRSLASELAASTTTSKS"
misc_feature    complement(28389. .30346)
                /gene="OSJNBa0089K24.6"
                /note="probably inactive due to frameshift(s) in CDS
                probably inactive due to no termination codon in CDS
                probably inactive due to stop codon(s) in CDS
                pseudogene, putative RIM2 protein"
                /pseudo
gene            complement(28389. .30346)
                /gene="OSJNBa0089K24.6"
misc_feature    37667. .39625
                /gene="OSJNBa0089K24.7"
                /note="probably inactive due to frameshift(s) in CDS
                probably inactive due to no termination codon in CDS
                probably inactive due to stop codon(s) in CDS
                pseudogene, putative RIM2 protein"
                /pseudo
gene            37667. .39625
                /gene="OSJNBa0089K24.7"
gene            join(42704. .42902,42996. .44002,44039. .44122,44225. .45342,
                45431. .45604,45694. .45804,45905. .45977,46073. .46171,
                46271. .46426,47025. .47183)
                /gene="OSJNBa0089K24.8"
CDS             join(42704. .42902,42996. .44002,44039. .44122,44225. .45342,
                45431. .45604,45694. .45804,45905. .45977,46073. .46171,
                46271. .46426,47025. .47183)
                /gene="OSJNBa0089K24.8"
                /note="hypothetical protein
                similar to Oryza sativa chromosome 1, P0489A01.2"
                /codon_start=1
                /protein_id="BAB62551.1"
                /db_xref="GI:15128223"
                /translation="MADRDEEQILYDTIALGSSQYWNKEEGNEDPNQYLNEEGNAERD
                AEGNEEGNVERDAEGNEEEASGNGRPSAPAQAAKNFVRHSGWVVRDNVPVSKVYWRRT
                RARGGDDSFVPESEKEMLWTTMLETFTLPAGTENIVKQWTLKKMAEQFQTFKGDLYRK
                YILKGQTPNFDIFPKLRDHWDEFIAYKTGQQGQAMMERNKENAAKKKYHHHLGSGGYS
                VAMPKWEEMEASLLERGIEPATAKWPDRSKFWYYAHGGMLNPVDGSLVFSDQIREAAN
                RLTDAVEASSQGTFRPDREKDELSLALQTPEHPGQTRGKGVIPWKMGFKEDIHTYRSR
                MRSKRDTEAKIADLEYRVSSYELSIQEEVARKVDERMAAHRSHDPQPYIPPPMVSPSG
                NRSSCASTGQDETTYPVDEITQRTPCELHIPFKNLSIKGMAIPTDISGTYHCRPIPAG
                YSRVEVELVEAVYEDIELDYPGGDGETHLRDTSHAIILWRKRYIIVPGRQAASRAPSP
                PAPPSPPAPSPPAPPPPPPALSPPAPPPPPPAPSPPAPLPPPPAPSPPAPPPPPPPPP
                PCPPAPPKTRSRQAPPPASTRATKKAKVDTTKNKEPPYGCSQEELDAYVAGEVKRQLK
                PQSPKKKIPINPSVKNFFKGMSTANKEALKLSDYDRTLRKAYYKKSKPVPQLGEQPHQ
                VVEPLVTGEDFGITDFISDIGLTMAQLVGGAPIPKAEVAYKFELGKPLVRPEQLQSLP
                TQMYKFHERYMEMSANGREMFGARIRNPDFLQGEDVLWIHFKDVFDLYHRDALDVSLL
                SAWILMEIQRARRRGVYDTGFIDPRKINTEMLDKYEKDTEDNLVHLLTQQHFKTFILL
                PYNTEFHWVLFFFDLDACRVTVYDSMNKEEKVFDKVFQLIDRAWDRFRQLVRGTWKEK
                LGRRFHFPCAKQDKGTNLCGYYVCEYAHCLSNQIYTKRELDRIHMREKLPHKDFITAV
                QEQLMGFINEEVLNPDGEFYYDGSTIRNVGPSSSDIVANEKMRRSQDLILNPSDNAVR
                SGRMRVGNGGNGIGPSRSHNRHRPLSQPC"
gene            join(52433. .52611,52978. .53083,54068. .54430)
                /gene="OSJNBa0089K24.9"
CDS             join(52433. .52611,52978. .53083,54068. .54430)
                /gene="OSJNBa0089K24.9"
                /note="hypothetical protein
                similar to Arabidopsis thaliana chromosome 1, F18B13.24"
                /codon_start=1
                /protein_id="BAB62552.1"
                /db_xref="GI:15128224"
                /translation="MVNTTAGVKCGGGGAALPLSTLNHVSLVCRSLSTSLTFYRDFLG
```

alignment_scores:

```
            Quality:   74.50            Length:    114
              Ratio:  1.242              Gaps:      5
 Percent Similarity:  52.632   Percent Identity: 24.561

alignment_block:
 US-10-048-197-2 x AP003215   ..

 Align seg 1/1  to: AP003215  from: 1  to: 154137

      13 LeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGlyAsnValAl 29
         ::::::      |||:::::::      ||||||      :::||||||
  147281 ATTATACTACGTTTAATACTTCAGATGTGTGTCCATATATCCAATGTGAT 147330

      29 aMetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLysGlyL 46
            |||   :::|||   :::   :::::::::::::::::::::::||||
  147331 AACCAAACCCAAAAACTTTTCCCAACTAAATAGGGCCTTAGTTTCGGGAA 147380

      46 ysThrAsnLysGlnGluIleSerSerArg.....PheGlySer 58
         |||||::::::   :::   :::   |||        ::::::
  147381 AGACGGACTCCCCTGACAGTAACATAAGAGAGAGAAGCGGGTTGTCTATG 147430

      59 AlaAspSer.....IleSerPheMetIleValValIleLysPheGl 72
         :::||||||        :::||||||   :::      :::::::::
  147431 TCGGACTCTCTTAGGTTAGTCTCTTTTTGTTTAGAAGGGATGGAGTATAC 147480

      72 yHisThr.....AlaIleLeuAlaP 79
            |||                 |||||||:::
  147481 CAGTACAGATCTTATTTATGTGCATCCCTCGCTATGACTAATCTTGTCTA 147530

      79 roAsn...ArgTrpGlnGluIleLeu...SerLeuIleIleSerPheLeu 93
           |||   ||||||      :::|||      |||           |||
  147531 CAAACCTACGGTGGATTGGAGTTCTGTAGCCCCTCCCCTGTCACGCACTG 147580

      94 TrpValLysProTyrArgProLysAsnLeuSerPheTyrLeu 107
         :::   |||:::   |||   :::|||      |||:::
  147581 AAACTTGTCCCCTTTTACCCCATCTCTTTGCTTGCATACATT 147622

seq_name: gb_htg:AC099643

seq_documentation_block:
LOCUS       AC099643              171660 bp    DNA     linear   HTG 16-NOV-2001
DEFINITION  Mus musculus clone RP24-350G1, WORKING DRAFT SEQUENCE, 35 unordered
            pieces.
ACCESSION   AC099643
VERSION     AC099643.1  GI:16948343
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 171660)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-350G1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171660)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
```

```
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L17460
                Center clone name: 350_G_1
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 161250 bases at least Q40
                Consensus quality: 165616 bases at least Q30
                Consensus quality: 167502 bases at least Q20
                Insert size: 188000; agarose-fp
                Insert size: 168260; sum-of-contigs
                Quality coverage: 12.7 in Q20 bases; agarose-fp
                Quality coverage: 14.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 35 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    78525: contig of 78525 bp in length
            *    78526 78625: gap of      100 bp
            *    78626    79185: contig of 560 bp in length
            *    79186 79285: gap of      100 bp
            *    79286    79495: contig of 210 bp in length
            *    79496 79595: gap of      100 bp
            *    79596    80537: contig of 942 bp in length
            *    80538 80637: gap of      100 bp
            *    80638    81632: contig of 995 bp in length
            *    81633 81732: gap of      100 bp
            *    81733    82472: contig of 740 bp in length
            *    82473 82572: gap of      100 bp
            *    82573    83248: contig of 676 bp in length
            *    83249 83348: gap of      100 bp
            *    83349    84207: contig of 859 bp in length
            *    84208 84307: gap of      100 bp
            *    84308    84514: contig of 207 bp in length
            *    84515 84614: gap of      100 bp
            *    84615    85295: contig of 681 bp in length
            *    85296 85395: gap of      100 bp
            *    85396    86100: contig of 705 bp in length
            *    86101 86200: gap of      100 bp
            *    86201    86869: contig of 669 bp in length
            *    86870 86969: gap of      100 bp
            *    86970    87738: contig of 769 bp in length
            *    87739 87838: gap of      100 bp
            *    87839    88571: contig of 733 bp in length
            *    88572 88671: gap of      100 bp
            *    88672    89751: contig of 1080 bp in length
            *    89752 89851: gap of      100 bp
            *    89852    90536: contig of 685 bp in length
            *    90537 90636: gap of      100 bp
            *    90637    91599: contig of 963 bp in length
            *    91600 91699: gap of      100 bp
            *    91700    92353: contig of 654 bp in length
            *    92354 92453: gap of      100 bp
            *    92454    93487: contig of 1034 bp in length
            *    93488 93587: gap of      100 bp
            *    93588    94231: contig of 644 bp in length
            *    94232 94331: gap of      100 bp
            *    94332    95444: contig of 1113 bp in length
            *    95445 95544: gap of      100 bp
            *    95545    96659: contig of 1115 bp in length
            *    96660 96759: gap of      100 bp
            *    96760    97953: contig of 1194 bp in length
            *    97954 98053: gap of      100 bp
            *    98054    98770: contig of 717 bp in length
            *    98771 98870: gap of      100 bp
            *    98871    99995: contig of 1125 bp in length
            *    99996 100095: gap of      100 bp
            *   100096   100895: contig of 800 bp in length
            *   100896 100995: gap of      100 bp
            *   100996   101656: contig of 661 bp in length
            *   101657 101756: gap of      100 bp
            *   101757   102779: contig of 1023 bp in length
            *   102780 102879: gap of      100 bp
            *   102880   104345: contig of 1466 bp in length
            *   104346 104445: gap of      100 bp
            *   104446   105396: contig of 951 bp in length
            *   105397 105496: gap of      100 bp
            *   105497   106713: contig of 1217 bp in length
            *   106714 106813: gap of      100 bp
            *   106814   108439: contig of 1626 bp in length
            *   108440 108539: gap of      100 bp
            *   108540   109727: contig of 1188 bp in length
            *   109728 109827: gap of      100 bp
            *   109828   148957: contig of 39130 bp in length
            *   148958 149057: gap of      100 bp
            *   149058   171660: contig of 22603 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171660
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP24-350G1"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1. .78525
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    78626. .79185
                     /note="assembly_fragment"
     misc_feature    79286. .79495
                     /note="assembly_fragment"
     misc_feature    79596. .80537
                     /note="assembly_fragment"
     misc_feature    80638. .81632
                     /note="assembly_fragment"
     misc_feature    81733. .82472
                     /note="assembly_fragment"
     misc_feature    82573. .83248
                     /note="assembly_fragment"
     misc_feature    83349. .84207
                     /note="assembly_fragment"
     misc_feature    84308. .84514
                     /note="assembly_fragment"
     misc_feature    84615. .85295
                     /note="assembly_fragment"
     misc_feature    85396. .86100
                     /note="assembly_fragment"
     misc_feature    86201. .86869
                     /note="assembly_fragment"
     misc_feature    86970. .87738
                     /note="assembly_fragment"
     misc_feature    87839. .88571
                     /note="assembly_fragment"
     misc_feature    88672. .89751
                     /note="assembly_fragment"
     misc_feature    89852. .90536
```

```
                /note="assembly_fragment"
     misc_feature    90637. .91599
                /note="assembly_fragment"
     misc_feature    91700. .92353
                /note="assembly_fragment"
     misc_feature    92454. .93487
                /note="assembly_fragment"
     misc_feature    93588. .94231
                /note="assembly_fragment"
     misc_feature    94332. .95444
                /note="assembly_fragment"
     misc_feature    95545. .96659
                /note="assembly_fragment"
     misc_feature    96760. .97953
                /note="assembly_fragment"
     misc_feature    98054. .98770
                /note="assembly_fragment"
     misc_feature    98871. .99995
                /note="assembly_fragment"
     misc_feature    100096. .100895
                /note="assembly_fragment"

alignment_scores:
            Quality:   74.50            Length:    121
              Ratio:  1.164              Gaps:      6
 Percent Similarity:  52.893   Percent Identity: 27.273

alignment_block:
 US-10-048-197-2 x AC099643/rev   ..

 Align seg 1/1  to reverse of: AC099643  from: 1  to: 171660

      22 CysValSerThrGlyAsnValAlaMetLys..................... 31
         |||::::::|||::::::   :::   |||
   26751 TGTCTTGAAACAAACAAACAAACAAACAAAACAAAACAAAACAAAACAAA 26702

      32 .........GluGlnAsnGlnGlnThrIleGluGlnThrIleIleLysG 45
                  :::::::|||   :::|||   :::|||||||      |||
   26701 ACAAAAAAACAAAAAAAACCCCAAAACATACAAACAAACAAAAAACAAAA 26652

      45 lyLysThrAsnLysGlnGluIleSerSerArgPheGly............ 57
         :::||||||||||:::::::::|||||||||||||
   26651 AACAAACAAACAAAAAAAACTTCAGTTCTAGATTCTGGAAGCCAGTGACT 26602

      58 ..........................SerAlaAspSerIleSerPheMe 65
                                  :::::::|||   |||   :::
   26601 TAAATTGAGGCCACAGGTGTTACTTGAGATACAGATTTTATTGTTTTGTG 26552

      65 tIleValValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgT 82
         :::|||:::   :::|||:::::::|||      |||:::   |||   |
   26551 GCTGGTGATCTCTTCCTTTAATAACACT......CTCTGTAGAAATACAT 26508

      82 rpGln..............GluIleLeuSerLeu.IleIleSerPhe.. 92
         ||                 |||:::   ||| |||:::::::|||
   26507 GGATATGCTTATAGTCTTGAATTATATTCATGTTGAATATTAACTTTCCC 26458

      93 .............LeuTrpValLysProTyrArgProLysAsnLeuSerP 105
                     |||   |||:::||||||   |||:::::::::::::
   26457 AGACAGACACAGTCTAGAAGTAAATCCCTATTGCCCTAGGGGATTTGACA 26408

     105 heTyrLeuThr 108
         :::   |||
   26407 CATGGAAGACA 26397

seq_name: gb_htg:AC021986

seq_documentation_block:
LOCUS       AC021986              172182 bp    DNA     linear   HTG 19-JUL-2000
DEFINITION  Homo sapiens clone RP11-4E13, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   AC021986
VERSION     AC021986.2  GI:9280771
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172182)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-4E13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172182)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 19, 2000 this sequence version replaced gi:6731251.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2818
                Center clone name: 4_E_13
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 164051 bases at least Q40
                Consensus quality: 168112 bases at least Q30
                Consensus quality: 169693 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 171282; sum-of-contigs
                Quality coverage: 4.7 in Q20 bases; agarose-fp
                Quality coverage: 5.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5011: contig of 5011 bp in length
            *     5012 5111: gap of      100 bp
            *     5112     9430: contig of 4319 bp in length
            *     9431 9530: gap of      100 bp
            *     9531    14192: contig of 4662 bp in length
            *    14193 14292: gap of      100 bp
            *    14293    20845: contig of 6553 bp in length
            *    20846 20945: gap of      100 bp
            *    20946    30131: contig of 9186 bp in length
            *    30132 30231: gap of      100 bp
            *    30232    42652: contig of 12421 bp in length
            *    42653 42752: gap of      100 bp
```

```
            *    42753    68827: contig of 26075 bp in length
            *    68828 68927: gap of      100 bp
            *    68928    98647: contig of 29720 bp in length
            *    98648 98747: gap of      100 bp
            *    98748   126143: contig of 27396 bp in length
            *   126144 126243: gap of      100 bp
            *   126244   172182: contig of 45939 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172182
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-4E13"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .5011
                     /note="assembly_fragment"
     misc_feature    5112. .9430
                     /note="assembly_fragment"
     misc_feature    9531. .14192
                     /note="assembly_fragment"
     misc_feature    14293. .20845
                     /note="assembly_fragment"
     misc_feature    20946. .30131
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    30232. .42652
                     /note="assembly_fragment"
     misc_feature    42753. .68827
                     /note="assembly_fragment"
     misc_feature    68928. .98647
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    98748. .126143
                     /note="assembly_fragment"
     misc_feature    126244. .172182
                     /note="assembly_fragment"
BASE COUNT    53360 a  30939 c  31936 g  55042 t    905 others
ORIGIN


alignment_scores:
           Quality:   74.50            Length:     85
             Ratio:  1.490              Gaps:      3
 Percent Similarity:  58.824   Percent Identity: 25.882

alignment_block:
 US-10-048-197-2 x AC021986/rev   ..

 Align seg 1/1  to reverse of: AC021986  from: 1  to: 172182

      22 CysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIl 38
         |||::::::||||||   |||:::   |||:::   :::|||::::::|| 
  134156 TGTGCTGCCACAGGGTTTGTGGTGTCAAAGCAGGCATCTCAGTCCAGCAT 134107

      38 eGlu.GlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSer 54
         ||||       ||||||             ::::::      :::     
  134106 TGAGGCCTTTCATCATCACAGTCTTTCTATCTGAAGTTTATCTT...... 134063

      55 ArgPheGlySerAlaAspSerIleSerPheMetIleValValIleLysPh 71
                              :::||||||:::|||:::      :::||  
  134062 .....................CTGTCATTCCTCATCATACCCACAAGATT 134034

      71 eGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerL 88
         ||||||||||:::::::::      :::           :::   :::   
  134033 TGGCCATACCTGTGTCATT......ACCTTTGTTGTTGTTGTACATACAC 133990

      88 euIleIleSerPheLeuTrpValLysProTyrArgProLysAsnLeuSer 104
           ::::::   ::::::|||:::   ||||||      ::::::|||:::  
  133989 ATGTTATGATTTACATTTGGGCTTTTCCTTATTCCTTTTCCACGTTAGCA 133940

     105 Phe 105
         |||
```

```
  133939 TTT 133937

seq_name: gb_pr:AC019184

seq_documentation_block:
LOCUS       AC019184              194372 bp    DNA     linear   PRI 09-MAY-2001
DEFINITION  Homo sapiens BAC clone RP11-298E6 from 2, complete sequence.
ACCESSION   AC019184
VERSION     AC019184.3  GI:14010913
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194372)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 194372)
  AUTHORS   Meyer,R., Maupin,R. and Bolla,S.
  TITLE     The sequence of Homo sapiens BAC clone RP11-298E6
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 194372)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 194372)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 9, 2001 this sequence version replaced gi:7109570.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0298E06
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
```

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-194L1; the clone sequenced to the right is RP11-442O5. Actual start of this clone is at base position 1 of RP11-298E6; actual end is at base position 194372 of RP11-298E6.

The sequence RP11-298E6 contains a dinucleotide (AT) repeat from base position 17640 to 17740 for which the exact length is unknown. Assembly of the database is consistent with digest information.

FEATURES	Location/Qualifiers
source	1. .194372
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-298E6"
	/clone_lib="RPCI-11"
repeat_region	453. .548
	/rpt_family="MIR"
misc_feature	704. .1261
	/note="similar to EST AI598155 (NID:g4607203) tn14c07.x1"
repeat_region	2991. .3284
	/rpt_family="Alu"
repeat_region	3436. .3559
	/rpt_family="Alu"
repeat_region	4128. .4261
	/rpt_family="MIR"
repeat_region	4444. .4725
	/rpt_family="L1"
repeat_region	4800. .5304
	/rpt_family="L1"
repeat_region	5341. .5639
	/rpt_family="Alu"
misc_feature	5883. .6120
	/note="similar to EST BE010465 (NID:g8270698)"
repeat_region	6707. .6786
	/rpt_family="MIR"
misc_feature	7844. .8165
	/note="similar to EST D29089 (NID:g599022)"
repeat_region	8389. .8504
	/rpt_family="MER94"
repeat_region	8850. .8975
	/rpt_family="MER1_type"
repeat_region	9838. .10131
	/rpt_family="Alu"
repeat_region	10325. .10602
	/rpt_family="L1"
repeat_region	10603. .10906
	/rpt_family="Alu"
repeat_region	10907. .11814
	/rpt_family="L1"
repeat_region	11923. .12053
	/rpt_family="L1"
repeat_region	12079. .12135
	/rpt_family="L1"
repeat_region	12479. .12733
	/rpt_family="L2"
repeat_region	13609. .13690
	/rpt_family="L1"
repeat_region	13730. .13985
	/rpt_family="L1"
repeat_region	14607. .14707
	/rpt_family="ERVL"
repeat_region	15168. .15464
	/rpt_family="Alu"
repeat_region	15527. .15872
	/rpt_family="L1"
repeat_region	15902. .16075
	/rpt_family="L1"
repeat_region	16584. .16880
	/rpt_family="Alu"
repeat_region	17081. .17221
	/rpt_family="MIR"
repeat_region	17311. .17598
	/rpt_family="Alu"
misc_feature	17910. .18238
	/note="similar to EST BE181265 (NID:g8660441)"
repeat_region	18393. .18521
	/rpt_family="L2"
repeat_region	18806. .19031
	/rpt_family="Alu"
repeat_region	19816. .19906
	/rpt_family="L2"
repeat_region	20001. .20075
	/rpt_family="MIR"
repeat_region	20884. .21192
	/rpt_family="Alu"
repeat_region	21352. .21655
	/rpt_family="Alu"
repeat_region	21930. .22238
	/rpt_family="Alu"
repeat_region	22775. .23299
	/rpt_family="MaLR"
repeat_region	23432. .23724
	/rpt_family="Alu"
repeat_region	24864. .25126
	/rpt_family="L1"
repeat_region	25127. .25431
	/rpt_family="Alu"
repeat_region	25432. .25482
	/rpt_family="L1"
repeat_region	25484. .25656
	/rpt_family="Alu"
repeat_region	25661. .25705
	/rpt_family="L1"
repeat_region	25706. .25982
	/rpt_family="Alu"
repeat_region	25983. .26468
	/rpt_family="L1"
repeat_region	26469. .26770
	/rpt_family="Alu"
repeat_region	26771. .26838
	/rpt_family="L1"
repeat_region	26872. .27607
	/rpt_family="L1"
repeat_region	27716. .27803
	/rpt_family="ERVL"
repeat_region	28014. .28147
	/rpt_family="MaLR"
repeat_region	28159. .28285
	/rpt_family="Alu"
repeat_region	28310. .28435
	/rpt_family="Alu"
repeat_region	28957. .29032
	/rpt_family="MER2_type"
repeat_region	29057. .29124
	/rpt_family="MER2_type"
repeat_region	30400. .30515
	/rpt_family="ERVL"
repeat_region	30575. .30924
	/rpt_family="MaLR"
repeat_region	30932. .31067

alignment_scores:
Quality: 74.50 Length: 85
Ratio: 1.490 Gaps: 3
Percent Similarity: 58.824 Percent Identity: 25.882

alignment_block:
US-10-048-197-2 x AC019184 ..

Align seg 1/1 to: AC019184 from: 1 to: 194372

```
      22 CysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIl 38
         |||::::::||||||   |||:::   |||:::   :::|||::::::||
   55291 TGTGCTGCCACAGGGTTTGTGGTGTCAAAGCAGGCATCTCAGTCCAGCAT 55340
```

```
     38 eGlu.GlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSer 54
        ||||       ||||||           ::::::      :::
  55341 TGAGGCCTTTCATCATCACAGTCTTTCTATCTGAAGTTTATCTT..... 55384

     55 ArgPheGlySerAlaAspSerIleSerPheMetIleValValIleLysPh 71
                          :::||||||:::|||:::      :::||
  55385 .....................CTGTCATTCCTCATCATACCCACAAGATT 55413

     71 eGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerL 88
        ||||||||||:::::::::      :::          :::   :::
  55414 TGGCCATACCTGTGTCATT......ACCTTTGTTGTTGTTGTACATACAC 55457

     88 euIleIleSerPheLeuTrpValLysProTyrArgProLysAsnLeuSer 104
          ::::::   :::::::|||:::   ||||||      ::::::|||:::
  55458 ATGTTATGATTTACATTTGGGCTTTTCCTTATTCCTTTTCCACGTTAGCA 55507

    105 Phe 105
        |||
  55508 TTT 55510
```

```
seq_name: gb_pr:AC062004

seq_documentation_block:
LOCUS       AC062004              205646 bp    DNA     linear   PRI 06-OCT-2001
DEFINITION  Homo sapiens chromosome 8, clone RP11-325O20, complete sequence.
ACCESSION   AC062004
VERSION     AC062004.5  GI:15982523
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 205646)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-325O20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 205646)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 205646)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 6, 2001 this sequence version replaced gi:14573732.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9902
                Center clone name: 325_O_20
            --------------.
FEATURES             Location/Qualifiers
     source          1. .205646
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-325O20"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   583. .613
                     /rpt_family="AT_rich"
     repeat_region   1040. .1082
                     /rpt_family="(T)n"
     repeat_region   complement(2328. .2465)
                     /rpt_family="HAL1b"
     repeat_region   complement(2521. .2963)
                     /rpt_family="MLT1C"
     repeat_region   3087. .3097
                     /rpt_family="AT_rich"
     repeat_region   complement(3098. .3387)
                     /rpt_family="AluSx"
     repeat_region   3388. .3410
                     /rpt_family="AT_rich"
     repeat_region   3524. .3573
                     /rpt_family="AT_rich"
     repeat_region   4752. .4793
                     /rpt_family="AT_rich"
     unsure          complement(5166. .5170)
                     /note="<30 qual SNGL region"
     repeat_region   5528. .5951
                     /rpt_family="MLT1A1"
     repeat_region   complement(6207. .6289)
                     /rpt_family="L1M4"
     repeat_region   6292. .6325
                     /rpt_family="MLT1E1"
     repeat_region   6326. .6360
                     /rpt_family="(CA)n"
     repeat_region   6361. .6846
                     /rpt_family="MLT1E1"
     repeat_region   complement(6996. .7320)
                     /rpt_family="MLT1A1"
     repeat_region   complement(7333. .8629)
                     /rpt_family="L1M4"
     repeat_region   9323. .9384
                     /rpt_family="(TA)n"
     repeat_region   complement(9459. .9638)
```

```
                /rpt_family="L1M4"
repeat_region   complement(9806. .10824)
                /rpt_family="L1MDa"
repeat_region   10826. .11001
                /rpt_family="(TA)n"
repeat_region   11009. .11167
                /rpt_family="(CATATA)n"
repeat_region   complement(11176. .11527)
                /rpt_family="L1MDa"
repeat_region   complement(11728. .12018)
                /rpt_family="L1MEc"
repeat_region   12123. .12159
                /rpt_family="AT_rich"
repeat_region   12423. .12734
                /rpt_family="AluSg"
repeat_region   12968. .13048
                /rpt_family="L2"
repeat_region   14102. .14245
                /rpt_family="MIR"
repeat_region   14322. .14359
                /rpt_family="AT_rich"
repeat_region   14370. .14543
                /rpt_family="L1ME3A"
repeat_region   14562. .14734
                /rpt_family="L1ME3A"
repeat_region   14749. .14788
                /rpt_family="AT_rich"
repeat_region   complement(14992. .15305)
                /rpt_family="AluY"
repeat_region   complement(15585. .15927)
                /rpt_family="MER7A"
repeat_region   15929. .15965
                /rpt_family="(TG)n"
repeat_region   16097. .16124
                /rpt_family="AT_rich"
repeat_region   16242. .16262
                /rpt_family="AT_rich"
repeat_region   16272. .16726
                /rpt_family="L1PA5"
repeat_region   16727. .17021
                /rpt_family="L1PA4"
repeat_region   17022. .22110
                /rpt_family="L1PA5"
repeat_region   22465. .22497
                /rpt_family="AT_rich"
repeat_region   22801. .23094
                /rpt_family="MLT1J2"
repeat_region   23609. .23630
                /rpt_family="AT_rich"
repeat_region   23688. .24063
                /rpt_family="MSTA"
repeat_region   24066. .25688
                /rpt_family="MSTA-int"
repeat_region   25689. .25753
                /rpt_family="MSTA"
repeat_region   complement(25751. .26310)
                /rpt_family="Tigger1"
repeat_region   26311. .26801
                /rpt_family="MER9"
repeat_region   complement(26802. .26822)
                /rpt_family="Tigger1"
repeat_region   complement(26952. .27720)
                /rpt_family="Tigger1"
repeat_region   27721. .27999
                /rpt_family="AluSc"
repeat_region   complement(28000. .28181)
                /rpt_family="Tigger1"
repeat_region   complement(28222. .29029)
                /rpt_family="L1MB8"
repeat_region   29037. .29105
                /rpt_family="(TA)n"
repeat_region   29865. .29894
                /rpt_family="AT_rich"
repeat_region   30308. .30347
                /rpt_family="(TA)n"
repeat_region   complement(31612. .31739)
                /rpt_family="MIR3"
repeat_region   32004. .32169

alignment_scores:
            Quality:   74.50              Length:     97
              Ratio:  1.380                Gaps:      5
 Percent Similarity:  55.670   Percent Identity: 28.866

alignment_block:
 US-10-048-197-2 x AC062004/rev   ..

 Align seg 1/1  to reverse of: AC062004  from: 1  to: 205646

      22 CysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGln..... 36
         |||   ||||||   |||      :::|||      |||:::|||
   46222 TGTTTTTCCACAGAAAAC......TTAAAGGTCCTAAACTCTCAATGGCC 46179

      37 ..............ThrIleGluGlnThrIleIleLysGlyLysThrA 48
                        |||||||||:::::      :::   :
   46178 AGTTCTTTCATTCTCCTATTCTGAGCAGACTCTAGTAATATTCAGATTCA 46129

      48 snLysGlnGluIleSerSerArgPheGlySerAlaAspSerIleSerPhe 64
         ::|||:::|||                |||:::   ::::::||||||
   46128 AAAAGAAAGAA..................AGCACTACTACTCTTTCTTTT 46097

      65 MetIleValValIleLysPheGlyHisThrAlaIleLeu...AlaProAs 80
         :::::::::      :::::::|||      ::::::      |||
   46096 TTTCTTCTTTTTTAAGAGATGGGGTCTCGCTCTGTTGCCCAGGCTGGAGT 46047

      80 nArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrpValLysP 97
          :::|||||||:::   |||:::|||:::               ::::::
   46046 GCAGTGGCAGGATTATCTCGCACTGCTTACTATA.........ACCTCAA 46006

      97 roTyrArgProLysAsnLeuSerPheTyrLeuThrAlaLys 110
           |||:::||||||   :::   :::   ||||||||||||
   46005 ACTACCAGCCTAAAGCGATCCTGTTACCTTTGACTGCCAAA 45965

seq_name: gb_ba:AF253043

seq_documentation_block:
LOCUS       AF253043               1377 bp    DNA     linear   BCT 03-NOV-2000
DEFINITION  Lactobacillus crispatus strain LMG 12003 surface layer protein
            (slpNA) gene, complete cds.
ACCESSION   AF253043
VERSION     AF253043.1  GI:7739699
KEYWORDS    .
SOURCE      Lactobacillus crispatus.
  ORGANISM  Lactobacillus crispatus
            Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
            Lactobacillus.
REFERENCE   1  (bases 1 to 1377)
  AUTHORS   Sillanpaa,J., Martinez,B., Antikainen,J., Toba,T., Kalkkinen,N.,
            Tankka,S., Lounatmaa,K., Keranen,J., Hook,M.,
            Westerlund-Wikstrom,B., Pouwels,P.H. and Korhonen,T.K.
  TITLE     Characterization of the collagen-binding S-layer protein CbsA of
            Lactobacillus crispatus
  JOURNAL   J. Bacteriol. 182 (22), 6440-6450 (2000)
  MEDLINE   20507814
   PUBMED   11053389
REFERENCE   2  (bases 1 to 1377)
  AUTHORS   Martinez,B., Smit,E. and Pouwels,P.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2000) IPLA-CSIC, Carretera de Infiesto s/n,
            Villaviciosa 33300, Spain
FEATURES             Location/Qualifiers
     source          1. .1377
                     /organism="Lactobacillus crispatus"
                     /strain="LMG 12003"
                     /db_xref="taxon:47770"
     gene            1. .1377
```

```
                     /gene="slpNA"
     CDS             1. .1377
                     /gene="slpNA"
                     /note="SlpNA"
                     /codon_start=1
                     /transl_table=11
                     /product="surface layer protein"
                     /protein_id="AAF68971.1"
                     /db_xref="GI:7739700"
                     /translation="MKKNLRIVSAAAAALLAVAPVAASAVSVNAADSTATTTANATDY
                     TNINLGGTTVSNTENQVDVTPSIVLNGNVKNTAGNLISKATLSGSISATFGGKSYTAN
                     LRGTDQNNVLINGRTAKDELSNVNAGSSNTITIKNVGFNFGPENKGKKITLVSSNSKV
                     TFGSDNAKTVTVSLDQNGTAKDLTVNISDVTAFNATNTNGVVFYNVTTGTQAHAGNAM
                     VLANTQGQLNTAALLPAIESNYVAVQRVDSDSANGNGTYNFADFKHVNNIEFATAIKD
                     QLKAQNIDVGPQGFFKAPHTFTVKVKATSSINGKSEELPVTFTVANVADPVVPSQPKT
                     IMHNAYYYKEDGTTRANNDKAKRYESVNVAMSTKKIGNKDFYEVIKDGKATGMYINAD
                     NIDGTKRTLKHNAYVYKTSKKRANKVVLKKGDTVVTYGGTYTFKNGKQYYKIYNNTEK
                     TYVKASNF"
BASE COUNT      445 a    276 c    261 g    395 t
ORIGIN


alignment_scores:
            Quality:   74.00            Length:     81
              Ratio:  1.451              Gaps:      3
 Percent Similarity:  62.963   Percent Identity: 27.160

alignment_block:
 US-10-048-197-2 x AF253043   ..

 Align seg 1/1  to: AF253043  from: 1  to: 1377

       1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
         :::||||||   :::::::::   |||   :::   |||:::::
     202 GTTCTTAATGGTAATGTCAAGAATACTGCTGGTAACCTTATTTCAAAGGC 251

      17 uLeuValThrGlyCysValSer.....................ThrG 26
             :::::::|||   :::|||                     |||:
     252 TACTTTGAGTGGTTCTATTTCAGCAACTTTCGGTGGCAAGAGCTACACTG 301

      26 lyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluGlnThrIle 42
         ::|||:::         :::||||||               :::::::::
     302 CTAACTTAAGAGGTACTGACCAAAAC.............AACGTTTTA 336

      43 IleLysGlyLysThrAsnLysGlnGluIleSerSer...ArgPheGlySe 58
         |||:::|||:::|||   |||:::|||:::|||:::   :::   |||||
     337 ATTAACGGTAGAACTGCTAAAGATGAATTAAGTAACGTTAATGCTGGCAG 386

      58 rAlaAspSerIleSerPheMetIleValValIleLysPheGly 72
         |:::::::::|||::::::      |||   ::::::||||||
     387 CTCCAACACTATTACTATCAAAAATGTTGGATTTAACTTTGGC 429

seq_name: gb_ba:AF421351

seq_documentation_block:
LOCUS       AF421351               2884 bp    DNA     linear   BCT 18-OCT-2001
DEFINITION  Azotobacter vinelandii L-isoaspartate o-methyltransferase (pcm)
            gene, partial cds; and murein endopeptidase (mep) and stationary
            phase sigma factor (rpoS) genes, complete cds.
ACCESSION   AF421351
VERSION     AF421351.1  GI:16226162
KEYWORDS    .
SOURCE      Azotobacter vinelandii.
  ORGANISM  Azotobacter vinelandii
            Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
            Azotobacter.
REFERENCE   1  (bases 1 to 2884)
  AUTHORS   Kujat Choy,S.L., Meakins,D., Tindale,A.E. and Page,W.J.
  TITLE     The stationary phase sigma factor (rpoS) of Azotobacter vinelandii
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2884)
  AUTHORS   Tindale,A.E., Meakins,D., Kujat Choy,S.L. and Page,W.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-2001) Biological Sciences, University of Alberta,
            CW405 Biological Sciences Bldg., Edmonton, Alta T6G 2E9, Canada
FEATURES             Location/Qualifiers
     source          1. .2884
                     /organism="Azotobacter vinelandii"
                     /strain="UW"
                     /db_xref="taxon:354"
     gene            <1. .396
                     /gene="pcm"
     CDS             <1. .396
                     /gene="pcm"
                     /codon_start=1
                     /transl_table=11
                     /product="L-isoaspartate o-methyltransferase"
                     /protein_id="AAL16089.1"
                     /db_xref="GI:16226163"
                     /translation="LEIGTGSGYQTAVLAQLVERVFSVERIQALQERAKERLLELNLR
                     NVVFRWGDGWEGWPALAPYNGIMVTAAASDVPSALLDQLAPGGRLVIPVGSGAEQQLL
                     LIVREPDGGFRRHALETVRFVPLLQGSLF"
     gene            1033. .1686
                     /gene="mep"
     CDS             1033. .1686
                     /gene="mep"
                     /note="similar to nlpD gene of Pseudomonas aeruginosa"
                     /codon_start=1
                     /transl_table=11
                     /product="murein endopeptidase"
                     /protein_id="AAL16090.1"
                     /db_xref="GI:16226164"
                     /translation="MRYGWEWRELAAHNGIAAPYVIHPGQKIQLGPVRQAAASTAARP
                     SVPAATPAPTVPATSQAAVAATPVKPDSGKVQPPPATKAPSAQPAVSARSPSGWAWPA
                     EGPLVGRFSSNGSLNKGIDIGGQLGQPVLAASDGSVVYAGGGLRGYGELIIIKHSDVY
                     VSAYGHNRRLLVREGQQVKAGQVIAEMGSTGTDRVKLHFEIRRQGKPVDPLQYLPSR"
     gene            1786. .2790
                     /gene="rpoS"
     CDS             1786. .2790
                     /gene="rpoS"
                     /note="transcriptional regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="stationary phase sigma factor"
                     /protein_id="AAL16091.1"
                     /db_xref="GI:16226165"
                     /translation="MALNIKAKEAPEFDVDDDVLLMDADIVLEDVLDDEPDLPVAKSR
                     QANLKQHKYIDYTRALDATQLYLNEIGFSPLLTPAEEVHFARLAQRGDPAGRKRMIES
                     NLRLVVKIARRYVNRGLSLLDLIEEGNLGLIRAVEKFDPERGFRFSTYATWWIRQTIE
                     RAIMNQTRTIRLPIHVVKELNVYLRAARELTQKLDHEPSAEEIANLLEKPVAEVKRML
                     GLNERVTSVDVSLGPDSDKTLLDVLTDDRPSDPCDLLQDDDLSQSIDQWLFDLTEKQR
                     EVVVRRFGLRGHESCTLEEVGQEIGLTRERVRQIQVEALKRLREILEKNGLSSDALFQ
                     "
BASE COUNT      509 a    846 c    959 g    570 t
ORIGIN


alignment_scores:
            Quality:   74.00            Length:     82
              Ratio:  1.510              Gaps:      3
 Percent Similarity:  59.756   Percent Identity: 26.829

alignment_block:
 US-10-048-197-2 x AF421351   ..

 Align seg 1/1  to: AF421351  from: 1  to: 2884

       6 IleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCy 22
         :::|||   ::::::::::::::::::   |||||||||:::::|||||
     874 GTTCGGAGCGTGCTGGGGGGGATCATCCTTGCCCTGCTGTTGGCCGGCTG 923

      22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleG 39
         |:::|||:::               ::::::::::|||   |||:::   |
     924 CGCCAGTTCC.............TCCCGGAACAACGTGCAGGTGGCGG 958

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
```

```
      ||:::     :::       |||:::  |||::::::  ||||||
   959 AGCGCCAGTCGGTTCCTCGCGTCACCAGCGGGCAACATGTGGTCAGCCGT 1008

    56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
         |||  :::|||||||:::             ::::::::||
  1009 GGCGAGTCGATCTCGTCCATTGCC.............ATGCGCTATGG 1043

    72 y.....................HisThrAlaIleLeuAlaPro 79
       |                     |||::::::|||  ||||||
  1044 CTGGGAGTGGCGCGAACTGGCCGCCCACAATGGCATCGCTGCGCCC 1089

seq_name: gb_htg:AC067965

seq_documentation_block:
LOCUS       AC067965              64165 bp    DNA     linear   HTG 15-JUN-2001
DEFINITION  Arabidopsis thaliana chromosome 1 clone T28N5, *** SEQUENCING IN
            PROGRESS ***, 11 unordered pieces.
ACCESSION   AC067965
VERSION     AC067965.2  GI:14456635
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 64165)
  AUTHORS   Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C.,
            Khan,S.; Kim,C., Altafi,H., Bei,Q., Chin,C., Chiou,J., Choi,E.,
            Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T.,
            Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
            Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J.,
            Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G.,
            Federspiel,N.A., Theologis,A. and Ecker,J.R.
  TITLE     Genomic sequence for Arabidopsis thaliana BAC T28N5 from chromosome
            I
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 64165)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
COMMENT     On Jun 15, 2001 this sequence version replaced gi:7656669.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4995: contig of 4995 bp in length
            *                   gap of unknown length
            *     4996    19478: contig of 14483 bp in length
            *                   gap of unknown length
            *    19479    21387: contig of 1909 bp in length
            *                   gap of unknown length
            *    21388    23278: contig of 1891 bp in length
            *                   gap of unknown length
            *    23279    24511: contig of 1233 bp in length
            *                   gap of unknown length
            *    24512    26159: contig of 1648 bp in length
            *                   gap of unknown length
            *    26160    29239: contig of 3080 bp in length
            *                   gap of unknown length
            *    29240    32881: contig of 3642 bp in length
            *                   gap of unknown length
            *    32882    37985: contig of 5104 bp in length
            *                   gap of unknown length
            *    37986    60939: contig of 22954 bp in length
            *                   gap of unknown length
            *    60940    64165: contig of 3226 bp in length.
FEATURES             Location/Qualifiers
     source          1. .64165
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="T28N5"
BASE COUNT    20083 a  12418 c  12567 g  19097 t
ORIGIN


alignment_scores:
            Quality:   74.00            Length:     67
              Ratio:  1.947              Gaps:      5
 Percent Similarity:  56.716   Percent Identity: 34.328

alignment_block:
 US-10-048-197-2 x AC067965   ..

 Align seg 1/1  to: AC067965  from: 1  to: 64165

     41 ThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPheGl 57
        :::|||   :::   |||||||||:::::::::
  58574 TCAATCAAACGGACCAAAACGAATCGCAAGAAA..................... 58606

     57 ySerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHisT 74
         |||   |||:::         |||:::::::::::::::|||||||||
  58607 .TCACTCGATCAAACG......ATGGTGATGCTCCTCTCATTTGGTCAT. 58648

     74 hrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIle 90
                  ||||||:::|||   :::|||||||   ||||||:::
  58649 .............CCAAATCAATGGCTCAAAATACTCCTCCTCATTTTA 58684

     91 ..................SerPheLeu...TrpValLysProTyrArgPr 100
                         |||||||||   |||   :::      |||||
  58685 CTCGATCAACAGTTGAGATCTTTCTTGACTTGGTTTGAAATTACTCGACC 58734

    100 o 100
        |
  58735 G 58735

seq_name: gb_htg:SPNEU1902

seq_documentation_block:
LOCUS       SPNEU1902             151947 bp    DNA     linear   HTG 11-JUL-2001
DEFINITION  Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***,
            in ordered pieces.
ACCESSION   AL449924
VERSION     AL449924.1  GI:11545149
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Streptococcus pneumoniae.
  ORGANISM  Streptococcus pneumoniae
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 151947)
  AUTHORS   Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,
            Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
            Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
            and Garcia-Bustos,J.F.
  TITLE     Annotated draft genomic sequence from a Streptococcus pneumoniae
            type 19F clinical isolate
  JOURNAL   Microb. Drug Resist. 7 (2), 99-125 (2001)
  MEDLINE   21335329
REFERENCE   2  (bases 1 to 151947)
  AUTHORS   Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
            Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
            Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
            Garcia-Bustos,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
            Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT     * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
```

```
FEATURES             Location/Qualifiers
     source          1. .151947
                     /organism="Streptococcus pneumoniae"
                     /serotype="19F"
                     /db_xref="taxon:1313"
                     /clone="G54"
BASE COUNT    45278 a  33410 c  27182 g  46003 t     74 others
ORIGIN


alignment_scores:
            Quality:   74.00              Length:    128
              Ratio:  1.121                Gaps:      7
 Percent Similarity:  51.562   Percent Identity: 29.688

alignment_block:
 US-10-048-197-2 x SPNEU1902/rev   ..

 Align seg 1/1  to reverse of: SPNEU1902  from: 1  to: 151947

      10 IleSerAlaLeuLeuThrAla...LeuLeuValThrGlyCysValSerTh 25
         ::::::||||||:::||||||   :::|||||||||||||   |||   ::
  119646 CTAGGAGCCTTGGTAACAGCTGTGATTCTCGTAACAGGCTCTGTTCTAGT 119597

      25 r......GlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
         :       ||||||:::         :::   |||   :::   ::::
  119596 CATTTTGGAAAATGTCACGAAGATTTTGCATCCGCAACCAGTCAATGATG 119547

      40 lnThrIleIle..................................... 43
         ::   |||:::
  119546 AGGGGATTCTCTGGTTAGGAATTATTGCGATTACTATCAATCTGTTAGCG 119497

      44 ..............LysGlyLysThrAsnLysGlnGluIleSerSerAr 55
                       ||||||||||||:::::::::::::|||   |||
  119496 AGTCTGGTGGTTGGTAAGGGAAAGACAAAGAATGAGTCTATTCTGAGTCT 119447

      55 gPheGlySerAlaAspSerIleSerPhe.............MetIleV 67
                     |||:::::::::::               |||   :
  119446 GCATTTTCTGGAAGATACGCTAGGGTGGGTAGCTGTTATCCTGATGGCGA 119397

      67 alValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 83
         ::|||::::::|||           ||||||   |||
  119396 TTGTTCTTCGATTTACGGACTGGTATATCCTAGATCCT............ 119359

      84 GluIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPr 100
         :::||||||||||:::||||||||||:::   ::::::
  119358 ...CTTTTGTCCCTTGTCATTTCTTTCTTTATTCTTTCA........... 119323

     100 oLysAsnLeuSer...PheTyrLeuThrAlaLys 110
          |||   ||||||   |||:::   |||   |||
  119322 .AAAGCCCTTTCACGTTTTTGGTCTACACTCAAG 119290

seq_name: gb_htg:AC009452

seq_documentation_block:
LOCUS       AC009452              174431 bp    DNA     linear   HTG 30-DEC-2001
DEFINITION  Homo sapiens chromosome 17 clone RP11-4F22 map 17, *** SEQUENCING
            IN PROGRESS ***, 3 unordered pieces.
ACCESSION   AC009452
VERSION     AC009452.17  GI:11560272
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174431)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-4F22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174431)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 6, 2000 this sequence version replaced gi:11276258.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L921
                Center clone name: 4_F_22
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    91566: contig of 91566 bp in length
            *    91567 91666: gap of      100 bp
            *    91667   121592: contig of 29926 bp in length
            *   121593 121692: gap of      100 bp
            *   121693   174431: contig of 52739 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174431
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-4F22"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    52366 a  39072 c  35991 g  46800 t    202 others
ORIGIN


alignment_scores:
            Quality:   74.00              Length:     90
              Ratio:  1.423                Gaps:       2
 Percent Similarity:  57.778   Percent Identity: 31.111

alignment_block:
 US-10-048-197-2 x AC009452   ..

 Align seg 1/1  to: AC009452  from: 1  to: 174431

      11 SerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGlyAs 27
         |||   ||||||      |||         :::|||::::::   |||::
   63447 TCCCAGCTACTCGGGAGACTGACACATGAGAACTGCTTGAACCCAGGAAG 63496

      27 nValAlaMetLysGlu.GlnAsnGlnGlnThrIleGluGlnThrIleIle 43
         :   :::   :::||| |||:::|||   |||:::   |||
   63497 CAGAGGTTGCAGTGAGCCAAGATCAAGCCACTGTGCTCCAGCCTGGACGA 63546

      44 LysGlyLysThr.....................AsnLysGlnGluIl 52
         ::::::||||||                     :::|||:::|||
```

```
   63547 CAGAGCAAGACCGTGTCTCAGAAAGAAAAAAAAAAAAAAAAAAAAGAATG 63596

      52 eSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValValI 69
         :::|||      :::   ::::::::::|||   :::   |||||||
   63597 CAACTCAGCACCCTGCACTTAGCAGACACTCAGTCAAGTGGCAGTTGTAA 63646

      69 leLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIle 85
         |||||   :::|||   :::      ||||||:::|||      ::::::
   63647 TTAAGTCATCACATCCCTCCCCTGCAGCTCCCCACAGACCAGGAAGCCTG 63696

      86 LeuSerLeuIleIleSer 91
         |||   |||   :::|||
   63697 CTCCCTTTAACAGTGTCG 63714
```

```
seq_name: gb_htg:AC099358

seq_documentation_block:
LOCUS       AC099358              177495 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus chromosome Rf4 clone CH230-34K4, *** SEQUENCING
            IN PROGRESS ***, 66 unordered pieces.
ACCESSION   AC099358
VERSION     AC099358.2  GI:17943937
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 177495)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177495)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:16901764.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GESX
                Center clone name: CH230-34K4
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 149775 bases at least Q40
                Consensus quality: 157121 bases at least Q30
                Consensus quality: 162139 bases at least Q20
                Estimated insert size: 143076; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 2.7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 66 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5095: contig of 5095 bp in length
            *     5096     5195: gap of unknown length
            *     5196    11383: contig of 6188 bp in length
            *    11384    11483: gap of unknown length
            *    11484    16683: contig of 5200 bp in length
            *    16684    16783: gap of unknown length
            *    16784    22608: contig of 5825 bp in length
            *    22609    22708: gap of unknown length
            *    22709    27938: contig of 5230 bp in length
            *    27939    28038: gap of unknown length
            *    28039    32330: contig of 4292 bp in length
            *    32331    32430: gap of unknown length
            *    32431    37713: contig of 5283 bp in length
            *    37714    37813: gap of unknown length
            *    37814    42192: contig of 4379 bp in length
            *    42193    42292: gap of unknown length
            *    42293    45821: contig of 3529 bp in length
            *    45822    45921: gap of unknown length
            *    45922    49329: contig of 3408 bp in length
            *    49330    49429: gap of unknown length
            *    49430    53383: contig of 3954 bp in length
            *    53384    53483: gap of unknown length
            *    53484    57060: contig of 3577 bp in length
            *    57061    57160: gap of unknown length
            *    57161    60495: contig of 3335 bp in length
            *    60496    60595: gap of unknown length
            *    60596    64032: contig of 3437 bp in length
            *    64033    64132: gap of unknown length
            *    64133    68040: contig of 3908 bp in length
            *    68041    68140: gap of unknown length
            *    68141    71756: contig of 3616 bp in length
            *    71757    71856: gap of unknown length
            *    71857    77090: contig of 5234 bp in length
            *    77091    77190: gap of unknown length
            *    77191    80409: contig of 3219 bp in length
            *    80410    80509: gap of unknown length
            *    80510    84845: contig of 4336 bp in length
            *    84846    84945: gap of unknown length
            *    84946    88506: contig of 3561 bp in length
            *    88507    88606: gap of unknown length
            *    88607    91717: contig of 3111 bp in length
```

```
*    91718    91817: gap of unknown length
*    91818    94912: contig of 3095 bp in length
*    94913    95012: gap of unknown length
*    95013    97740: contig of 2728 bp in length
*    97741    97840: gap of unknown length
*    97841   101192: contig of 3352 bp in length
*   101193   101292: gap of unknown length
*   101293   103818: contig of 2526 bp in length
*   103819   103918: gap of unknown length
*   103919   107092: contig of 3174 bp in length
*   107093   107192: gap of unknown length
*   107193   109752: contig of 2560 bp in length
*   109753   109852: gap of unknown length
*   109853   112708: contig of 2856 bp in length
*   112709   112808: gap of unknown length
*   112809   114829: contig of 2021 bp in length
*   114830   114929: gap of unknown length
*   114930   117858: contig of 2929 bp in length
*   117859   117958: gap of unknown length
*   117959   121771: contig of 3813 bp in length
*   121772   121871: gap of unknown length
*   121872   124073: contig of 2202 bp in length
*   124074   124173: gap of unknown length
*   124174   126622: contig of 2449 bp in length
*   126623   126722: gap of unknown length
*   126723   128819: contig of 2097 bp in length
*   128820   128919: gap of unknown length
*   128920   130969: contig of 2050 bp in length
*   130970   131069: gap of unknown length
*   131070   133523: contig of 2454 bp in length
*   133524   133623: gap of unknown length
*   133624   135654: contig of 2031 bp in length
*   135655   135754: gap of unknown length
*   135755   137129: contig of 1375 bp in length
*   137130   137229: gap of unknown length
*   137230   138685: contig of 1456 bp in length
*   138686   138785: gap of unknown length
*   138786   140784: contig of 1999 bp in length
*   140785   140884: gap of unknown length
*   140885   142169: contig of 1285 bp in length
*   142170   142269: gap of unknown length
*   142270   143518: contig of 1249 bp in length
*   143519   143618: gap of unknown length
*   143619   145099: contig of 1481 bp in length
*   145100   145199: gap of unknown length
*   145200   146491: contig of 1292 bp in length
*   146492   146591: gap of unknown length
*   146592   147869: contig of 1278 bp in length
*   147870   147969: gap of unknown length
*   147970   149003: contig of 1034 bp in length
*   149004   149103: gap of unknown length
*   149104   151134: contig of 2031 bp in length
*   151135   151234: gap of unknown length
*   151235   152932: contig of 1698 bp in length
*   152933   153032: gap of unknown length
*   153033   154184: contig of 1152 bp in length
*   154185   154284: gap of unknown length
*   154285   155403: contig of 1119 bp in length
*   155404   155503: gap of unknown length
*   155504   156702: contig of 1199 bp in length
*   156703   156802: gap of unknown length
*   156803   158164: contig of 1362 bp in length
*   158165   158264: gap of unknown length
*   158265   159542: contig of 1278 bp in length
*   159543   159642: gap of unknown length
*   159643   160720: contig of 1078 bp in length
*   160721   160820: gap of unknown length
*   160821   161992: contig of 1172 bp in length
*   161993   162092: gap of unknown length
*   162093   163584: contig of 1492 bp in length
*   163585   163684: gap of unknown length

alignment_scores:
```

```
            Quality:   74.00              Length:     72
              Ratio:  1.644                Gaps:      3
 Percent Similarity:  62.500   Percent Identity: 29.167

alignment_block:
 US-10-048-197-2 x AC099358/rev   ..

 Align seg 1/1  to reverse of: AC099358  from: 1  to: 177495

       5 HisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGl 21
         |||:::::::::|||:::|||   :::|||      |||      |||||
  119124 CACGTCCAGGTCACGGTGTCCAGGGTGTTGGGGTACCTAAGCAGTACGGG 119075

      21 yCysValSerThrGlyAsnValAlaMetLysGluGlnAsn...GlnGlnT 37
         ||||      ||||||:::   :::   |||      :::   :::   |
  119074 TTGC......ACAGGGACACCTGGGGAGAAGGCTCCTAGTTTAAAAGTGA 119031

      37 hrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSer 53
         ||   ::::::   ::::::   :::   ::::::|||   :::|||
  119030 CAAGACAGGAGCGGTTGGTGCACAATCCTTCTGGAAAAATCAGAATCTGT 118981

      54 SerArgPhe.........GlySerAlaAspSerIleSerPheMetIleVa 67
         :::::::|||           :::   :::   |||||||||   :::||
  118980 GGCCACTTCATTTTAGATTTCACCCGCGAGAGGATCTCATTCCGGGTAGT 118931

      67 lValIleLysPheGly 72
         |   ::::::||||||
  118930 CTTTCTAGAATTCGGG 118915

seq_name: gb_pr:AL137020

seq_documentation_block:
LOCUS       AL137020              182231 bp    DNA     linear   PRI 03-JAN-2002
DEFINITION  Human DNA sequence from clone RP11-146B14 on chromosome 9. Contains
            part of the gene KIAA0780, ESTs, STSs and GSSs, complete sequence.
ACCESSION   AL137020
VERSION     AL137020.13  GI:10086020
KEYWORDS    HTG; KIAA0780.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182231)
  AUTHORS   Babbage,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 11, 2000 this sequence version replaced gi:10039482.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            This sequence is the entire insert of clone RP11-146B14 The true
            left end of clone RP11-169L18 is at 151570 in this sequence. The
            true right end of clone RP11-403H13 is at 88099 in this sequence.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            RP11-146B14 is from the library RPCI-11.1 constructed by the group
```

```
          of Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .182231
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-146B14"
                     /clone_lib="RPCI-11.1"
     misc_feature    7. .689
                     /note="match: GSS: Em:AQ375850"
     repeat_region   365. .649
                     /note="AluSx repeat: matches 1. .286 of consensus"
     misc_feature    2945. .3711
                     /note="match: GSS: Em:AQ745052"
     repeat_region   3339. .3449
                     /note="L2 repeat: matches 2577. .2704 of consensus"
     repeat_region   3655. .3710
                     /note="28 copies 2 mer at 71% conserved"
     repeat_region   4627. .4758
                     /note="FLAM_C repeat: matches 1. .130 of consensus"
     repeat_region   5487. .5608
                     /note="MIR repeat: matches 60. .173 of consensus"
     repeat_region   5775. .6071
                     /note="AluSx repeat: matches 1. .299 of consensus"
     repeat_region   6886. .6984
                     /note="MIR repeat: matches 48. .136 of consensus"
     repeat_region   7180. .7468
                     /note="AluSx repeat: matches 7. .299 of consensus"
     repeat_region   8427. .8541
                     /note="AluJo repeat: matches 22. .136 of consensus"
     repeat_region   9507. .9649
                     /note="MIR repeat: matches 48. .188 of consensus"
     repeat_region   10006. .10255
                     /note="MLT1-INTERNAL repeat: matches 630. .878 of
                     consensus"
     gene            join(12340. .12533,15596. .15834,17774. .18096,21847. .21955,
                     42825. .43006,44915. .45128,46976. .47052,77894. .77949,
                     80124. .80232,134726. .134911,159107. .159277)
                     /gene="bA146B14.1"
     CDS             join(<12340. .12533,15596. .15834,17774. .18096,21847. .21955,
                     42825. .43006,44915. .45128,46976. .47052,77894. .77949,
                     80124. .80232,134726. .134911,159107. .>159277)
                     /gene="bA146B14.1"
                     /note="match: cDNAs: Em:AB018323
                     match: ESTs: Em:AA361338 Em:AW385519 Em:AW372902
                     Em:AI969877 Em:AW449099 Em:AW402136 Em:T53845
                     match: proteins: Tr:O94877"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bA146B14.1 (KIAA0780)"
                     /protein_id="CAD20146.1"
                     /db_xref="GI:18070851"
                     /translation="CTCRKDMVKISMDIFVRKFQPDRYQLWKQGKDIYTIDHTKPTPA
                     STPEVKAWLQRRRKVRKASRSFQCARSTSKRPKADEEEEVSDEVDGAEVPNPDSVTDD
                     LKVSEKSEAAVKLRNTEASSEEESSASRMQVEQNLSDHIKLSGNSCLSTSVTEDIKTE
                     DDKAYAYRSVPSISSEADDSIPLSSGYEKPEKSDPSELSWPKSPESCSSVAESNGVLT
                     EGEESDVESHGNGLEPGEIPAVPSGERNSFKVPSIAEGENKTSKSWRHPLSRPPARSP
                     MTLVKQQAPSDEELPEVLSIEEEVEETESWAKPLIHLWQTKSPNFAAEQEYNATVARM
                     KPHCAICTLLMPYHKPDSSNEENDARWETKLDEVVTSEGKTKPLIPEMCFIYSEENIE
                     YSPPNAFLEEDGTSLLISCAKCCVRVHASCYGIPSHEICDGWLCARCKRNAWTAECCL
                     CNLRGGALKQTKNNKWAHVMCAVAVPEVRFTNVPERTQIDVGRIPLQRLKLKCIFCRH
                     RVKRVSGACIQCSYGRCPASFHVTCAHAAGVLMEPDDWPYVVNITCFRHKVNPNVKSK
                     ACEKVISVGQTVITKHRNTRYYSCRVMAVTSQTFYEVMFDDGSFSRDTFPEDIV"
     repeat_region   14870. .14986
                     /note="FLAM_C repeat: matches 1. .118 of consensus"
     repeat_region   14989. .15042
                     /note="27 copies 2 mer ac 100% conserved"
     repeat_region   15214. .15303
                     /note="45 copies 2 mer tt 66% conserved"
     repeat_region   16435. .16776
                     /note="L2 repeat: matches 1907. .2289 of consensus"
     repeat_region   17112. .17420
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   17636. .17681
                     /note="23 copies 2 mer tg 78% conserved"
     repeat_region   18345. .18526
                     /note="L1MD1 repeat: matches 6036. .6219 of consensus"
     misc_feature    18595. .19112
                     /gene="bA146B14.1"
                     /note="match: GSS: Em:AQ548118"
     repeat_region   19305. .19618
                     /note="AluJo repeat: matches 1. .312 of consensus"
     misc_feature    complement(19955. .20234)
                     /note="match: STS: Em:G10183"
     misc_feature    19974. .20146
                     /gene="bA146B14.1"
                     /note="match: STS: Em:G15769"
     repeat_region   20136. .20513
                     /note="L2 repeat: matches 2310. .2718 of consensus"
     repeat_region   21215. .21508
                     /note="AluJo repeat: matches 1. .292 of consensus"
     repeat_region   22510. .22813
                     /note="AluJb repeat: matches 3. .297 of consensus"
     repeat_region   23184. .23313
                     /note="FLAM_A repeat: matches 1. .129 of consensus"
     repeat_region   25395. .25719
                     /note="MER1B repeat: matches 1. .331 of consensus"
     repeat_region   26758. .26797
                     /note="20 copies 2 mer ac 90% conserved"
     repeat_region   27107. .27399
                     /note="AluY repeat: matches 1. .293 of consensus"
     misc_feature    complement(27309. .27637)
                     /note="match: GSS: Em:AQ043656"
     repeat_region   27422. .27556
                     /note="L1PB3 repeat: matches 6010. .6141 of consensus"
     misc_feature    complement(28428. .28893)
                     /note="match: GSS: Em:AQ691327"
     repeat_region   29079. .29160
                     /note="MIR repeat: matches 53. .142 of consensus"
     repeat_region   29947. .30240
                     /note="AluSx repeat: matches 1. .294 of consensus"
     repeat_region   32144. .32184
                     /note="MIR repeat: matches 70. .111 of consensus"
     repeat_region   33021. .33313
                     /note="AluSx repeat: matches 7. .299 of consensus"
     repeat_region   33452. .33627
                     /note="L1MA7 repeat: matches 6093. .6279 of consensus"
     repeat_region   33871. .34002
                     /note="MIR repeat: matches 8. .140 of consensus"
     repeat_region   34003. .34292
                     /note="AluSx repeat: matches 1. .285 of consensus"
     repeat_region   36003. .36280
                     /note="LTR22 repeat: matches 188. .509 of consensus"
     repeat_region   36281. .36579
                     /note="AluSq repeat: matches 1. .300 of consensus"
     repeat_region   36580. .36780
                     /note="LTR22 repeat: matches 1. .188 of consensus"
     repeat_region   38781. .39023
                     /note="Charlie1 repeat: matches 2522. .2761 of consensus"
     repeat_region   39032. .40398
                     /note="L1M3e repeat: matches -878. .1358 of consensus"
     repeat_region   40400. .40895
                     /note="Charlie1a repeat: matches 4. .548 of consensus"
     repeat_region   40913. .41316
                     /note="TIGGER2 repeat: matches 1. .450 of consensus"
     repeat_region   41889. .42199
                     /note="AluSq repeat: matches 1. .311 of consensus"
     repeat_region   43198. .43468
                     /note="AluJo repeat: matches 38. .298 of consensus"
     misc_feature    complement(45910. .46385)
                     /note="match: GSS: Em:AQ608877"
     repeat_region   46043. .46076
                     /note="17 copies 2 mer ac 100% conserved"
     misc_feature    46392. .46807
```

```
                /gene="bA146B14.1"
                /note="match: GSS: Em:AQ175358"
     repeat_region   47247. .47549
                /note="AluY repeat: matches 1. .303 of consensus"
     repeat_region   47550. .47848
                /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   48060. .48358
                /note="AluJo repeat: matches 2. .299 of consensus"
     repeat_region   49173. .49484
                /note="MER7A repeat: matches 3. .346 of consensus"
     repeat_region   51706. .52084
                /note="L1MA2 repeat: matches 5937. .6305 of consensus"
     misc_feature    52075. .52416

alignment_scores:
            Quality:   74.00            Length:     99
              Ratio:  1.396              Gaps:      5
 Percent Similarity:  53.535   Percent Identity: 29.293

alignment_block:
 US-10-048-197-2 x AL137020   ..

 Align seg 1/1  to: AL137020  from: 1  to: 182231

      22 CysValSerThr......GlyAsnValAlaMetLysGluGlnAsnGlnGl 36
         |||||||:::|||      |||:::|||:::            :::|||::
  166688 TGTGTAAACACAGATGTAGGAGGAGTGACATGGATGTTCTGGAGGCAAAA 166737

      36 nThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleS 53
         :||||||   :::   :::      |||||||||:::|||:::   |||
  166738 TACAATACCCAAGGAAGTT......GGAAAGACTAAGAAAAAAGCCATTC 166781

      53 erSerArgPheGlySerAlaAspSerIleSerPheMetIleValValIle 69
            :::|||           :::|||:::   |||:::
  166782 CCAGAAACTTT..............TTCAGCTATGGGATCTTGGGAAGC 166816

      70 LysPheGlyHisThrAlaIleLeuAla..................... 78
         :::|||||||   :::::::::   :::
  166817 AATTTTGGATCTTCGTCGGTGTCTACACTTTGAGGTGCTCTGGAGTTTTG 166866

      79 .........ProAsnArgTrpGlnGluIleLeu.SerLeuIleIleSer 91
                 |||::::::||||||      ||| |||:::      |||
  166867 GGCCAAGTGGCCAAGACAGTGGCAGACATCTCTCATCAGTGAGCCCATCT 166916

      92 PheLeuTrpValLysProTyrArgProLysAsnLeuSerPheTyr 106
            ||||||           :::|||:::   |||   :::|||
  166917 CAGCTTTGGGAAGTCACATCCCATCCCAGCCCACTGATGTATTAT 166961

seq_name: gb_htg:AC099386

seq_documentation_block:
LOCUS       AC099386              217121 bp    DNA     linear   HTG 21-DEC-2001
DEFINITION  Rattus norvegicus chromosome Rf4 clone CH230-163F6, WORKING DRAFT
            SEQUENCE, 35 unordered pieces.
ACCESSION   AC099386
VERSION     AC099386.4  GI:17974833
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 217121)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 217121)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 21, 2001 this sequence version replaced gi:17062973.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
              Center project name: GKEA
                Center clone name: CH230-163F6
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 199000 bases at least Q40
                Consensus quality: 202571 bases at least Q30
                Consensus quality: 205231 bases at least Q20
                Estimated insert size: 199252; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 35 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17030: contig of 17030 bp in length
            *    17031    17130: gap of unknown length
            *    17131    34262: contig of 17132 bp in length
            *    34263    34362: gap of unknown length
            *    34363    49178: contig of 14816 bp in length
            *    49179    49278: gap of unknown length
            *    49279    64140: contig of 14862 bp in length
```

```
            *    64141    64240: gap of unknown length
            *    64241    74691: contig of 10451 bp in length
            *    74692    74791: gap of unknown length
            *    74792    84936: contig of 10145 bp in length
            *    84937    85036: gap of unknown length
            *    85037    94557: contig of 9521 bp in length
            *    94558    94657: gap of unknown length
            *    94658   105308: contig of 10651 bp in length
            *   105309   105408: gap of unknown length
            *   105409   114496: contig of 9088 bp in length
            *   114497   114596: gap of unknown length
            *   114597   121368: contig of 6772 bp in length
            *   121369   121468: gap of unknown length
            *   121469   129890: contig of 8422 bp in length
            *   129891   129990: gap of unknown length
            *   129991   136269: contig of 6279 bp in length
            *   136270   136369: gap of unknown length
            *   136370   142620: contig of 6251 bp in length
            *   142621   142720: gap of unknown length
            *   142721   149174: contig of 6454 bp in length
            *   149175   149274: gap of unknown length
            *   149275   154421: contig of 5147 bp in length
            *   154422   154521: gap of unknown length
            *   154522   161138: contig of 6617 bp in length
            *   161139   161238: gap of unknown length
            *   161239   168686: contig of 7448 bp in length
            *   168687   168786: gap of unknown length
            *   168787   172245: contig of 3459 bp in length
            *   172246   172345: gap of unknown length
            *   172346   178286: contig of 5941 bp in length
            *   178287   178386: gap of unknown length
            *   178387   182921: contig of 4535 bp in length
            *   182922   183021: gap of unknown length
            *   183022   186678: contig of 3657 bp in length
            *   186679   186778: gap of unknown length
            *   186779   190709: contig of 3931 bp in length
            *   190710   190809: gap of unknown length
            *   190810   193501: contig of 2692 bp in length
            *   193502   193601: gap of unknown length
            *   193602   196269: contig of 2668 bp in length
            *   196270   196369: gap of unknown length
            *   196370   199716: contig of 3347 bp in length
            *   199717   199816: gap of unknown length
            *   199817   202043: contig of 2227 bp in length
            *   202044   202143: gap of unknown length
            *   202144   204384: contig of 2241 bp in length
            *   204385   204484: gap of unknown length
            *   204485   206937: contig of 2453 bp in length
            *   206938   207037: gap of unknown length
            *   207038   208913: contig of 1876 bp in length
            *   208914   209013: gap of unknown length
            *   209014   210723: contig of 1710 bp in length
            *   210724   210823: gap of unknown length
            *   210824   211969: contig of 1146 bp in length
            *   211970   212069: gap of unknown length
            *   212070   213625: contig of 1556 bp in length
            *   213626   213725: gap of unknown length
            *   213726   214893: contig of 1168 bp in length
            *   214894   214993: gap of unknown length
            *   214994   216014: contig of 1021 bp in length
            *   216015   216114: gap of unknown length
            *   216115   217121: contig of 1007 bp in length.
FEATURES             Location/Qualifiers
     source          1. .217121
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /chromosome="Rf4"
                     /clone="CH230-163F6"
BASE COUNT    57695 a  47818 c  48957 g  59178 t   3473 others
ORIGIN


alignment_scores:
                Quality:   74.00              Length:      72
                  Ratio:  1.644                Gaps:       3
  Percent Similarity:  62.500    Percent Identity: 29.167

alignment_block:
 US-10-048-197-2 x AC099386/rev   ..

 Align seg 1/1  to reverse of: AC099386  from: 1  to: 217121

       5 HisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGl 21
         |||:::::::::|||:::|||   :::|||      |||      |||||
  131463 CACGTCCAGGTCACGGTGTCCAGGGTGTTGGGGTACCTAAGCAGTACGGG 131414

      21 yCysValSerThrGlyAsnValAlaMetLysGluGlnAsn...GlnGlnT 37
         ||||      ||||||:::   :::   |||      :::   :::   |
  131413 TTGC......ACAGGGACACCTGGGGAGAAGGCTCCTAGTTTAAAAGTGA 131370

      37 hrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSer 53
         ||   ::::::   ::::::   :::   ::::::|||   :::|||
  131369 CAAGACAGGAGCGGTTGGTGCACAATCCTTCTGGAAAAATCAGAATCTGT 131320

      54 SerArgPhe.........GlySerAlaAspSerIleSerPheMetIleVa 67
         ::::::|||            :::   :::   |||||||||   :::||
  131319 GGCCACTTCATTTTAGATTTCACCCGCGAGAGGATCTCATTCCGGGTAGT 131270

      67 lValIleLysPheGly 72
         |   ::::::||||||
  131269 CTTTCTAGAATTCGGG 131254
```

THIS PAGE BLANK (USPTO)

```
OM of: US-10-048-197-2 to: EST:*   out_format : pfs

Date: Sep 18, 2002  5:13 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US10048197/runat_17092002_135849_22587/app_query.fasta_1.168
 -DB=EST -QFMT=fastap -SUFFIX=std.rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500
 -MINLEN=0 -MAXLEN=2000000000 -USER=US10048197_@CGN1_1_4116
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-10-048-197-2
Query length: 111
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1805.460000


score_list:
Sequence          Strd Orig    ZScore   EScore Len  ! Documentation ..
gb_gss:AZ533627     -   84.00  175.34  1.61      906  ! AZ533627 ENTBU30TF Entamoeba hi
gb_gss:BH150303     -   84.00  175.13  1.65      923  ! BH150303 ENTPI29TR Entamoeba hi
gb_est1:AW352637    -   82.50  176.05  1.47      617  ! AW352637 660033D06.x1 660 - Mix
gb_est1:BB840173    -   81.00  175.77  1.52      459  ! BB840173 BB840173 RIKEN full-le
gb_est2:BE575116    -   76.50  165.17  5.93      457  ! BE575116 946087A11.x2 946 - tas
gb_est1:AI833467    -   76.50  163.18  7.65      547  ! AI833467 605089G04.x1 605 - End
gb_est2:T41696      +   76.50  162.03  8.87      607  ! T41696 10277 Lambda-PRL2 Arabid
gb_est2:BM412909    +   76.50  159.76  11.87     745  ! BM412909 EST587236 tomato break
gb_est1:AI068415    +   76.50  157.44  15.98     918  ! AI068415 mgae0002bA08f Magnapor
gb_est1:AW933759    -   76.00  162.20  8.68      537  ! AW933759 EST359602 tomato fruit
gb_est2:BI679657    +   75.50  165.15  5.95      370  ! BI679657 949031A01.y3 949 - Juv
gb_est1:AW017581    -   75.50  164.85  6.18      380  ! AW017581 614060H04.x1 614 - roo
gb_est2:BI698181    +   75.50  162.98  7.86      450  ! BI698181 949031A01.y4 949 - Juv
gb_est1:AW018175    -   75.50  162.76  8.08      459  ! AW018175 614068C08.x1 614 - roo
gb_est1:AI834558    -   75.50  162.73  8.11      460  ! AI834558 606069H03.x1 606 - Ear
gb_est2:BI245316    +   75.50  162.57  8.28      467  ! BI245316 949031A01.x2 949 - Juv
gb_est2:BE511166    -   75.50  162.08  8.82      488  ! BE511166 946058D02.x1 946 - tas
gb_est2:BI431074    -   75.50  162.01  8.89      491  ! BI431074 949069B08.x1 949 - Juv
gb_est1:AW037170    +   75.50  161.99  8.92      492  ! AW037170 614018H02.y1 614 - roo
gb_est2:BI478440    +   75.50  161.55  9.44      512  ! BI478440 949069B08.y1 949 - Juv
gb_est1:AW171836    -   75.50  161.52  9.46      513  ! AW171836 618051B11.x1 618 - Inh
gb_est1:AI586595    -   75.50  161.50  9.49      514  ! AI586595 486049C06.x4 486 - lea
gb_est2:BI643532    -   75.50  161.42  9.60      518  ! BI643532 949079C10.x1 949 - Juv
gb_est1:AI629856    -   75.50  161.18  9.89      529  ! AI629856 486041B10.x2 486 - lea
gb_est2:BI180391    -   75.50  161.10  9.99      533  ! BI180391 949031A01.x1 949 - Juv
gb_est1:AW090901    +   75.50  160.67  10.56     554  ! AW090901 614068C08.y1 614 - roo
gb_est2:BI675165    -   75.50  160.61  10.64     557  ! BI675165 949079C10.x2 949 - Juv
gb_est1:AI714378    -   75.50  160.57  10.69     559  ! AI714378 606013A04.x2 606 - Ear
gb_est1:AW566286    +   75.50  160.49  10.80     563  ! AW566286 660069A03.y1 660 - Mix
gb_est2:BF728138    -   75.50  160.03  11.46     587  ! BF728138 1000056E09.x2 1000 - U
gb_est2:BG842695    -   75.50  159.81  11.80     599  ! BG842695 MEST39-A03.T3 ISUM4-TN
gb_est1:AI691931    -   75.50  159.75  11.88     602  ! AI691931 606007A02.x1 606 - Ear
gb_est2:BM079588    -   75.50  159.66  12.02     607  ! BM079588 MEST96-F07.T3 ISUM4-TN
gb_est1:AW324655    -   75.50  159.59  12.13     611  ! AW324655 707036H01.x1 707 - Mix
gb_est1:AI711637    -   75.50  159.55  12.19     613  ! AI711637 605058E06.x1 605 - End
gb_est1:AW927882    +   75.50  159.53  12.22     614  ! AW927882 945010D06.Y1 945 - Mix
gb_est1:AI820102    -   75.50  159.35  12.50     624  ! AI820102 605086A01.x3 605 - End
gb_est2:BE509707    -   75.50  159.18  12.79     634  ! BE509707 946020H06.x1 946 - tas
gb_est1:AI665346    -   75.50  159.11  12.90     638  ! AI665346 605010B03.x1 605 - End
gb_est1:AW313326    -   75.50  158.90  13.25     650  ! AW313326 707024C08.x1 707 - Mix
gb_est2:BG841347    -   75.50  157.31  16.24     750  ! BG841347 MEST20-E05.T3 ISUM4
gb_est1:AU138543    +   75.50  156.74  17.48     790  ! AU138543 AU138543 PLACE1 Hom
gb_est2:BI389512    -   74.50  162.40  8.46      383  ! BI389512 949062C09.x1 949 -
gb_est2:BG992931    +   74.50  157.91  15.04     574  ! BG992931 MR3-HT0999-050201-0
gb_est1:AI692100    -   74.50  157.87  15.12     576  ! AI692100 606012C12.x1 606 -

seq_name: gb_gss:AZ533627

seq_documentation_block:
LOCUS       AZ533627               906 bp    DNA     linear   GSS 03-NOV-2000
DEFINITION  ENTBU30TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ533627
VERSION     AZ533627.1  GI:11089466
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 906)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 17
            High quality sequence stop: 773.
FEATURES             Location/Qualifiers
     source          1. .906
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      318 a    100 c    108 g    380 t
ORIGIN


alignment_scores:
            Quality:   84.00             Length:     92
              Ratio:  1.556               Gaps:      2
 Percent Similarity:  58.696   Percent Identity: 28.261

alignment_block:
 US-10-048-197-2 x AZ533627/rev   ..

 Align seg 1/1  to reverse of: AZ533627  from: 1  to: 906

      11 SerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGlyAs 27
         |||   ||||||   :::   |||   |||||||||
     621 TCAAATTTATTATTCACAACACTTCAAACTGGTTGT.............. 586
```

```
      27 nValAlaMetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleL 44
                 :::   |||:::|||:::               |||:::|||:
     585 .......TTATTAGAACGAAATGAAATAAAAGAAATAGGTACATTAATAG 543

      44 ysGlyLysThrAsnLysGlnGluIleSerSerArgPheGlySerAlaAsp 60
         ::   |||   ::::::::::::::::|||||:::|||   :::   |||
     542 AA...AAAATCTCAAATTCAGACTTATCTTCAAAATTCACAGAGCAAGAT 496

      61 SerIleSerPheMetIleValValIleLysPheGlyHisThrAlaIleLe 77
            |||   :::      |||:::|||:::::::::::::         ||
     495 GTTATTTGTATTGGAAGAGTTATTATTAGTATGTCTGAATGTTATACATT 446

      77 uAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuT 94
         |      :::   |||:::   :::||||||:::::|||:::::
     445 ATTAAAACAAGCTTGGGAACCTCTATTAAGTATTGTAATAAATATTAAAA 396

      94 rpValLysProTyrArgProLysAsn 102
         ::::::      |||      |||
     395 TTATAAATCAAGCAAGGGATATTAAT 370

seq_name: gb_gss:BH150303

seq_documentation_block:
LOCUS       BH150303                923 bp    DNA     linear   GSS 27-AUG-2001
DEFINITION  ENTPI29TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH150303
VERSION     BH150303.1  GI:15311742
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 923)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 19
            High quality sequence stop: 759.
FEATURES             Location/Qualifiers
     source          1. .923
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      307 a    126 c    108 g    382 t
ORIGIN

alignment_scores:
            Quality:   84.00            Length:     92
              Ratio:  1.556              Gaps:      2
 Percent Similarity:  58.696   Percent Identity: 28.261

alignment_block:
 US-10-048-197-2 x BH150303/rev   ..

 Align seg 1/1  to reverse of: BH150303  from: 1  to: 923

      11 SerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGlyAs 27
         |||   ||||||   :::   |||   |||||||||
     308 TCAAATTTATTATTCACAACACTTCAAACTGGTTGT.............. 273

      27 nValAlaMetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleL 44
                 :::   |||:::|||:::               |||:::|||:
     272 .......TTATTAGAACGAAATGAAATAAAAGAAATAGGTACATTAATAG 230

      44 ysGlyLysThrAsnLysGlnGluIleSerSerArgPheGlySerAlaAsp 60
         ::   |||   ::::::::::::::::|||||:::|||   :::   |||
     229 AA...AAAATCTCAAATTCAGACTTATCTTCAAAATTCACAGAGCAAGAT 183

      61 SerIleSerPheMetIleValValIleLysPheGlyHisThrAlaIleLe 77
            |||   :::      |||:::|||:::::::::::::         ||
     182 GTTATTTGTATTGGAAGAGTTATTATTAGTATGTCTGAATGTTATACATT 133

      77 uAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuT 94
         |      :::   |||:::   :::||||||:::::|||:::::
     132 ATTAAAACAAGCTTGGGAACCTCTATTAAGTATTGTAATAAATATTAAAA 83

      94 rpValLysProTyrArgProLysAsn 102
         ::::::      |||      |||
      82 TTATAAATCAAGCAAGGGATATTAAT 57

seq_name: gb_est1:AW352637

seq_documentation_block:
LOCUS       AW352637                617 bp    mRNA    linear   EST 02-FEB-2000
DEFINITION  660033D06.x1 660 - Mixed stages of anther and pollen Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW352637
VERSION     AW352637.1  GI:6851627
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 660033  row: D  column: 06.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="660 - Mixed stages of anther and pollen"
                     /tissue_type="whole premieotic anthers to pollen shed"
                     /dev_stage="premieotic anthers to pollen shed"
                     /lab_host="XLOLR"
                     /note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
                     Site_2: XhoI; Anther and pollen cDNA library.
                     Directionally sequenced with 5' end at the EcoRI site.
```

```
                    Created by Amie Franklin."
BASE COUNT      173 a    148 c    123 g    173 t
ORIGIN


alignment_scores:
            Quality:   82.50             Length:    101
              Ratio:  1.587               Gaps:      4
 Percent Similarity:  51.485   Percent Identity: 25.743

alignment_block:
 US-10-048-197-2 x AW352637/rev   ..

 Align seg 1/1  to reverse of: AW352637  from: 1  to: 617

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     379 AAGGCTACAGTCAAGTTTCAACTGAAGAAAGTCCTTTGCATGGGT...GT 333

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||::::::|||:::|||   |||   ||||||:::|||:
     332 CGCGGTCGGTAACCTGTCTATGGAGGAGAAGCAGATCCAACAGAACATCC 283

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     282 AGATG..................................................... 278

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                       ||||||:::|||::::::   ::::::|||
     277 ..............AGCATCAACTTCCTCGTGTCTCTTCTGAAG..... 248

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                            ::::::||||||::::::      |||
     247 .....................AAGAACTGGCAGAATGTAAGATGTCTCT 220

      89 leIleSerPheLeuTrpValLysProTyrArgProLysAsnLeuSerPhe 105
           |||:::            ||||||:::|||   |||::::::   |||
     219 ACATCAAGAGCACCATGGGGAAACCATTCAGGATTAAAAAAATAATGTTT 170

     106 Tyr 106
         :::
     169 TTT 167

seq_name: gb_est1:BB840173

seq_documentation_block:
LOCUS       BB840173                459 bp    mRNA    linear   EST 21-NOV-2001
DEFINITION  BB840173 RIKEN full-length enriched, 12 days embryo whole body Mus
            musculus cDNA clone E970005N21 5', mRNA sequence.
ACCESSION   BB840173
VERSION     BB840173.1  GI:17040904
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .459
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="E970005N21"
                     /clone_lib="RIKEN full-length enriched, 12 days embryo
                     whole body"
                     /tissue_type="whole body"
                     /dev_stage="12 days embryo"
BASE COUNT      126 a     70 c    126 g    137 t
ORIGIN


alignment_scores:
            Quality:   81.00             Length:    137
              Ratio:  1.191               Gaps:      4
 Percent Similarity:  49.635   Percent Identity: 22.628

alignment_block:
 US-10-048-197-2 x BB840173/rev   ..

 Align seg 1/1  to reverse of: BB840173  from: 1  to: 459

       4 HisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeu......Le 18
         |||   :::   |||      :::   :::||||||:::|||      ||
     455 CATACTTTCTTGTTGCTCTATAAGGATTTTCTGACCTGTCTTGACTGCCT 406

      18 uValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnG 35
         |   |||       ::::::::::::|||   ::::::|||||||||:
     405 CTACACACAGTTTCAAACAAGTAACAGAGTCCATATATCAGAACAAAACA 356

      35 lnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGlu 51
         ::          :::         ||||||:::|||::::::::::::
     355 AGACCAAAACTATCAGACCTACCCACAAAGGGTCAACGGGAGAAAAAAAC 306

      52 IleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValVa 68
         |||:::::::::   ||||||   |||      :::   :::|||:::||
     305 ATAACTACACAGCGAGGATCTATTGATCACTACGCTGAGGTTATCATAGT 256

      68 lIle..................................................... 69
         |:::
     255 TCTTAACTTGTCCCCAAAACCAGGAACACTTAACTTGTCCCAAAAGCCAA 206

      70 .....................................LysPhe 71
                                              |||
     205 GGAACACTTAACTTGTCCCAAAACCGGGAACATTAACTTGTCCCAAAGCC 156

      72 GlyHisThrAlaIleLeu.AlaProAsnArgTrpGlnGluIleLeuSerL 88
```

```
      |||:::   ::::::|||  :::|||      ||||||            |
    155 GGGAACCTATCGTTACTTTGTGCCTGCTTCCTGGCAA...........C 118

     88 euIleIleSerPheLeuTrpValLysProTyrArgProLysAsnLeuSer 104
        ||   |||   :::   :::::::::   :::   ||||||   |||:::
    117 TTTATATTCCCCTCTATTTTATCCGAGGTCACTTCCCAAAACTCCTGGGT 68

    105 PheTyrLeu 107
        :::::::::
     67 TACTTTGTG 59
```

seq_name: gb_est2:BE575116

```
seq_documentation_block:
LOCUS       BE575116               457 bp    mRNA    linear   EST 15-AUG-2000
DEFINITION  946087A11.x2 946 - tassel primordium prepared by Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   BE575116
VERSION     BE575116.1  GI:9824809
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 946087  row: A  column: 11.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Zea mays"
                     /cultivar="OH43"
                     /db_xref="taxon:4577"
                     /clone_lib="946 - tassel primordium prepared by Schmidt
                     lab"
                     /tissue_type="tassels"
                     /dev_stage="just after the transition from vegetative to
                     inflorescence development"
                     /lab_host="XLOLR"
                     /note="Organ: tassels; Vector: HybriZAP; Site_1: EcoRI;
                     Site_2: XhoI; George Chuck dissected immature tassels
                     between 1mm and 3mm.  Sharon  Stanfield prepared the cDNA
                     library in HybriZAP.  Sample insert size  range was 350 bp
                     to 3 Kb with a 1 Kb average."
BASE COUNT      114 a    118 c    107 g    118 t
ORIGIN
```

```
alignment_scores:
           Quality:   76.50            Length:     94
             Ratio:  1.628              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BE575116/rev   ..

 Align seg 1/1  to reverse of: BE575116  from: 1  to: 457

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     288 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 242

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
```

```
      |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
    241 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC.192

     39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
        :::::
    191 AGATG..................................................... 187

     56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                   |||:::::::|||::::::   :::::::|||
    186 ..............AGCGTGAACTTCCTTGTTTCCATTATGAAG..... 157

     72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                             ::::::||||||::::::      |||
    156 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 129

     89 leIleSerPheLeuTrpValLysProTyrArg 99
          ::::::            ||||||||||||
    128 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 97
```

seq_name: gb_est1:AI833467

```
seq_documentation_block:
LOCUS       AI833467               547 bp    mRNA    linear   EST 02-FEB-2000
DEFINITION  605089G04.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
            cDNA, mRNA sequence.
ACCESSION   AI833467
VERSION     AI833467.1  GI:5455777
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 605089  row: G  column: 04.
FEATURES             Location/Qualifiers
     source          1. .547
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
                     /tissue_type="nucellar, embryo, and endosperm"
                     /dev_stage="10-14 days post-pollination"
                     /lab_host="DH5(alpha)"
                     /note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI;
                     Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
                     lab"
BASE COUNT      152 a    132 c    123 g    140 t
ORIGIN
```

```
alignment_scores:
           Quality:   76.50            Length:     94
             Ratio:  1.628              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AI833467/rev   ..

 Align seg 1/1  to reverse of: AI833467  from: 1  to: 547

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
```

```
      :::   |||::::::      |||      :::|||      |||   ||
    386 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 340

     23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
        |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
    339 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 290

     39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
        :::::
    289 AGATG..................................................... 285

     56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   :::::::|||
    284 ..............AGCGTGAACTTCCTTGTTTTCCTTTTGAAG..... 255

     72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                             ::::::||||||:::::      |||
    254 .....................AAGAACTGGCAGAATGTGAGATGCCTTT 227

     89 leIleSerPheLeuTrpValLysProTyrArg 99
          ::::::            ||||||||||||
    226 ACGTCAAGAGCACCATGGGGAAACCATACAGG 195
```

seq_name: gb_est2:T41696

seq_documentation_block:

LOCUS T41696 607 bp mRNA linear EST 07-AUG-1995
DEFINITION 10277 Lambda-PRL2 Arabidopsis thaliana cDNA clone 64B8T7, mRNA sequence.
ACCESSION T41696
VERSION T41696.1 GI:931332
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 607)
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel,E. and Somerville,C.
TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL Plant Physiol. 106, 1241-1255 (1994)
MEDLINE 95148729
COMMENT On Jan 24, 1995 this sequence version replaced gi:634284.
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E. Lansing,Mi
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@ibm.cl.msu.edu
Seq primer: T7 dye primer.
FEATURES Location/Qualifiers
source 1. .607
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/clone="64B8T7"
/clone_lib="Lambda-PRL2"
/note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques. The vector is BRL's lambda Zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dT primed cDNA. "
BASE COUNT 139 a 128 c 146 g 164 t 30 others

ORIGIN

alignment_scores:
Quality: 76.50 Length: 95
Ratio: 1.319 Gaps: 5
Percent Similarity: 61.053 Percent Identity: 25.263

alignment_block:
US-10-048-197-2 x T41696 ..

Align seg 1/1 to: T41696 from: 1 to: 607

```
      3 AsnHisHisIleArgLeuThrIleSerAlaLeuLeuThr..... 15
        |||::::::::::   :::|||:::||||||      |||
    176 AACAGAAGAGTATCGGTTACCGTTTCCGCCGCCGCAACAACTGAACCNNT 225

     16 .AlaLeuLeuValThrGlyCysValSer...ThrGlyAsnValAla.... 29
         ::::::|||||||||||||:::   :::   ||||||::::::::::
    226 CACCGTCCTCGTTACCGGTGCCGNTNGAAGAACAGGACAAATTGTGTATA 275

     30 .....MetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLys 44
           :::||||||::::::::|||   :::   :::   :::::::::
    276 AGAAATTGAAGGAGAGGTCGGAGCAGTTCGTGGCAAGGGGTTTAGTGAGG 325

     45 GlyLysThrAsnLysGlnGluIleSerSerArgPheGlySer........ 58
        |||   :::|||::::::|||::::::::       |||
    326 ACAAAGGAGAGTAAGGAGAAGATCAATGGAGAGGATGNAAGTATTCATTT 375

     59 ..........AlaAspSerIleSerPheMetIleValValIleLys.... 70
                 :::|||:::   |||      ::::::::::::::
    376 GGNGNTATTAGGNGATACTGCATCAANTTNGCTCCTGCTNGTTGAAGGGG 425

     71 ...........PheGlyHisThrAlaIleLeuAla 78
                  ||||||            ||||||
    426 NTTTNANTGCTTTTGGGTCAATTCTTACCTTAGCG 460
```

seq_name: gb_est2:BM412909

seq_documentation_block:

LOCUS BM412909 745 bp mRNA linear EST 22-JAN-2002
DEFINITION EST587236 tomato breaker fruit Lycopersicon esculentum cDNA clone cLEG61H17 5' end, mRNA sequence.
ACCESSION BM412909
VERSION BM412909.1 GI:18264539
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 745)
AUTHORS Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato fruit tissue, breaker stage (2002)
JOURNAL Unpublished (2002)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
This clone is available through the Clemson University Genomics Institute
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .745
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEG61H17"
/clone_lib="tomato breaker fruit"

```
                     /tissue_type="Pericarp"
                     /dev_stage="breaker"
                     /lab_host="SOLR"
                     /note="Vector: pBluescriptSKmCUadapt; Site_1: EcoR1;
                     Site_2: XhoI; supplier: Boyce Thompson Institute;
                     sequencing: The Institute for Genomic Research.  Fruit
                     were harvested at the breaker stage (first sign of
                     lycopene accumulation on the blossom end of fruit).  Fruit
                     were cut in half and the seeds and locules were discarded
                     prior to freezing the pericarp."
BASE COUNT      211 a    123 c    178 g    233 t
ORIGIN


alignment_scores:
            Quality:   76.50             Length:     89
              Ratio:  1.443               Gaps:      4
 Percent Similarity:  59.551   Percent Identity: 29.213

alignment_block:
 US-10-048-197-2 x BM412909   ..

 Align seg 1/1  to: BM412909  from: 1  to: 745

      10 IleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGl 26
         :::||||||         ::::::   |||:::   |||      :::
      36 GTTTCTGCGCCAAGTAAATCTGTGAAGGTTTCTTGTTGTGAAAGAGTGCA 85

      26 yAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleI 43
         :::::::::   |||   |||                        |||   :
      86 GAGATTGGTATGTAAAGCACAA.................GCAACTGATG 117

      43 leLysGlyLysThrAsnLysGlnGluIleSerSerArgPheGlySerAla 59
         :::::   |||   :::|||:::|||            ||||||:::|||
     118 TTAATCGTAAAGAGACCAAGAATGAA.........TCTTTTGGTGAGGCT 158

      60 ...AspSerIleSerPheMetIleValValIleLysPheGlyHisThrAl 75
            |||:::::::::            |||:::||||||||   :::::
     159 CCTGATGAATTGACA.........TGTGTTATGAAGTTTGGTGGTTCATC 199

      75 aIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIleSerP 92
         ::::   :::   :::|||   :::|||:::   :::||||||:::||||
     200 AGTAGCCTCAGCCGAAAGAATGAGAGAAGTAGCTGATCTTATTCTTAGCT 249

      92 heLeuTrpValLysPro 97
         |||||      :::|||
     250 TTCTAGAAGAGAGACCT 266

seq_name: gb_est1:AI068415

seq_documentation_block:
LOCUS       AI068415                 918 bp    mRNA    linear   EST 09-DEC-1999
DEFINITION  mgae0002bA08f Magnaporthe grisea Appressorium Stage cDNA Library
            Magnaporthe grisea cDNA clone mgae0002bA08f 5', mRNA sequence.
ACCESSION   AI068415
VERSION     AI068415.1  GI:3391390
KEYWORDS    EST.
SOURCE      Magnaporthe grisea.
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 918)
  AUTHORS   Choi,W., Fang,E., Sasinowski,M., Wing,R. and Dean,R.A.
  TITLE     Expressed sequence characterization during appressorium formation
            in rice blast fungus, Magnaporthe grisea
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Dean,R.A.
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson Universiy, Clemson, SC 29634, USA
            Tel: 864 656 5737
            Fax: 864 656 4293
            Email: rdean@clemson.edu
            Seq primer: T3 primer (AATTAACCCTCACTAAAGGG)
            High quality sequence stop: 331.
FEATURES             Location/Qualifiers
     source          1. .918
                     /organism="Magnaporthe grisea"
                     /strain="70-15"
                     /db_xref="taxon:148305"
                     /clone="mgae0002bA08f"
                     /clone_lib="Magnaporthe grisea Appressorium Stage cDNA
                     Library"
                     /dev_stage="Germinated conidia on appressorium-inductive
                     surface"
                     /note="Vector: pBlueScriptII SK(+) Vector; Site_1: EcoRI;
                     Site_2: XhoI; The appressorium formation-specific cDNA
                     library was constructed  form conidia germinated for 5-8
                     hr on an inductive surface.   The library has an average
                     insert size of 1.5 kbp."
BASE COUNT      241 a    237 c    227 g    201 t     12 others
ORIGIN


alignment_scores:
            Quality:   76.50             Length:    113
              Ratio:  1.594               Gaps:      3
 Percent Similarity:  42.478   Percent Identity: 22.124

alignment_block:
 US-10-048-197-2 x AI068415   ..

 Align seg 1/1  to: AI068415  from: 1  to: 918

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||   ||||||::::::::::::|||:::            |||   ||
     107 AGAACCACCATAACCACTGTCGTGACTACAACCAATAGCGATGGCTCGGT 156

      23 lSerThr..................................... 25
         |||||||
     157 TTCCACCCAGACCAGCCAGATTGTCCAGGAAACTACTGTGCCCAGTTCCA 206

      26 ........GlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGlu 39
                    |||   :::|||   |||:::::::::::||||||
     207 ACAACAACGAGAACAACGGTATGGACGAGAAGCAGCGCAACACTATCATC 256

      40 GlnThrIleIleLysGly........................... 45
            :::::::::   |||
     257 GGAGCTGTTGTCGGTGGTGTCGGTGGTGCCATCGTGCTGGCTGTCATTGG 306

      46 ............................LysThrAsnLysGlnGluIleS 53
                                     |||   |||::::::|||::::
     307 TCTGGTTGCTTGGCGTGTATGGTGTCGCAAGAAGAACCAGGAGGAGGTGG 356

      53 erSerArgPheGlySerAlaAspSerIleSerPheMetIleValValIle 69
         :::::   ||||||::::::   |||
     357 ACAACTCTTTTGGAACCTCTGCCTCA...................... 382

      70 LysPheGlyHisThrAlaIleLeuAlaProAsnArgTrp 82
                    |||::::::   :::||||||||||||
     383 ............ACGTCGCTGAGAAGCCCGAACCGGTGG 409

seq_name: gb_est1:AW933759

seq_documentation_block:
LOCUS       AW933759                 537 bp    mRNA    linear   EST 18-MAY-2001
DEFINITION  EST359602 tomato fruit mature green, TAMU Lycopersicon esculentum
            cDNA clone cLEF56018 5', mRNA sequence.
ACCESSION   AW933759
VERSION     AW933759.1  GI:8109160
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
```

```
            Lycopersicon.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
            Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.
            , Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato fruit tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .537
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEF56018"
                     /clone_lib="tomato fruit mature green, TAMU"
                     /tissue_type="fruit pericarp"
                     /dev_stage="mature green (3-5 days pre-ripening)"
                     /lab_host="SOLR"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; cLEF - Fruit were tagged at the 1cm stage and
                     harvested 3-5 days prior to  ripening. Fruit were cut in
                     half to verify the seeds were indeed  'immature' and the
                     seeds and locules were discarded prior to freezing the
                     pericarp"
BASE COUNT      160 a     89 c    114 g    174 t
ORIGIN


alignment_scores:
            Quality:   76.00             Length:     79
              Ratio:  1.520               Gaps:      3
 Percent Similarity:  63.291   Percent Identity: 26.582

alignment_block:
 US-10-048-197-2 x AW933759/rev   ..

 Align seg 1/1  to reverse of: AW933759  from: 1  to: 537

      40 GlnThrIleIleLys...GlyLysThrAsnLysGlnGluIleSerSerAr 55
         |||   :::::::|||      :::   |||   :::|||:::   :::||
     479 CAACCGCTACTGAAATGTCATCGTGGAAATGTCGACGAACTCCACGATCG 430

      55 gPheGlySerAlaAspSerIleSerPhe.....MetIleValV 68
         ||||:::|||:::   :::   ||||||          :::::::::|
     429 ATTTTCTTCAAGTCTGAATACCTCATTTCCCTTTTCTTTGCTGCTTCTTG 380

      68 alIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnGlu 84
         ||:::::|||:::   |||      |||:::|||   :::|||::::::
     379 TAGTGCAGTTTTCACTAACCTTCTAGCTATCCCCTGTCAAGTGGCGAAAG 330

      85 ..............IleLeuSerLeuIleIleSerPheLeuTrpValLy 96
                   :::::::::|||::::::   |||      :::||
     329 TAGCCTCTTTAAGTTTTGATCCAACTGTTTTTTTTTTTCCAACTTCTCAA 280

      96 sProTyrArgProLysAsnLeuSerPheTyrLeuThr 108
         |   :::|||         |||   :::|||::::::
     279 GGAGTTTAGAAACACATAGCTTTTATACTATATATCA 243

seq_name: gb_est2:BI679657

seq_documentation_block:
LOCUS       BI679657                370 bp    mRNA    linear   EST 17-SEP-2001
DEFINITION  949031A01.y3 949 - Juvenile leaf and shoot cDNA from Steve Moose
            Zea mays cDNA, mRNA sequence.
ACCESSION   BI679657
VERSION     BI679657.1  GI:15632564
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 370)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 949031  row: A  column: 01.
FEATURES             Location/Qualifiers
     source          1. .370
                     /organism="Zea mays"
                     /cultivar="W64A"
                     /db_xref="taxon:4577"
                     /clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
                     Moose"
                     /tissue_type="immature leaf primordium and vegetative
                     meristem"
                     /dev_stage="4 stages from 3-13 days after imbibing"
                     /lab_host="E. coli XLOLR"
                     /note="Organ: juvenile vegetative shoots; Vector:
                     pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
                     of total RNA by weight from 4 tissue sources (see below)
                     were pooled, polyA+ RNA isolated, and cDNA synthesized for
                     EcoRI (5') and XhoI (3') directional cloning into lambda
                     Hybrizap vector from Stratagene. Tissue Sources: 1. Whole
                     shoots 3 days after sowing/imbibing in wet soil. 2. Basal
                     1.5 cm shoots 6 days after sowing - includes yellow
                     portions of developing leaves 1-5, primordia from 6-8, and
                     the vegetative apex. 3. Non-green portions of developing
                     leaves 4-5 and the vegetative apex,  including adult leaf
                     primordia, 9 days after sowing. 4. Partially expanded and
                     greening leaves 4-5 at 13 days after sowing."
BASE COUNT       91 a     97 c     97 g     85 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BI679657   ..

 Align seg 1/1  to: BI679657  from: 1  to: 370

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     130 AAGGCAACCGTCAAATTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 176

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |:::::||||||:::||||||:::|||   |||   ||||||:::||||:
     177 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 226

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     227 AGATG..................................................... 231

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                       |||:::::::|||::::::   ::::::|||
     232 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 261

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                            ::::::||||||::::::      |||
```

```
    262 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 289

     89 leIleSerPheLeuTrpValLysProTyrArg 99
        ::::::            ||||||||||||
    290 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 321

seq_name: gb_est1:AW017581

seq_documentation_block:
LOCUS       AW017581               380 bp    mRNA    linear   EST 13-SEP-1999
DEFINITION  614060H04.x1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW017581
VERSION     AW017581.1  GI:5871140
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614060  row: H  column: 04.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      103 a    101 c     80 g     96 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AW017581/rev   ..

 Align seg 1/1  to reverse of: AW017581  from: 1  to: 380

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     353 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 307

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |:::::|||||:::||||||:::|||   |||   ||||||:::|||:
     306 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 257

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     256 AGATG..................................................... 252

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                      |||::::::|||::::::   :::::::|||
```

```
    251 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 222

     72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                               ::::::||||||::::::      |||
    221 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 194

     89 leIleSerPheLeuTrpValLysProTyrArg 99
        ::::::            ||||||||||||
    193 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 162

seq_name: gb_est2:BI698181

seq_documentation_block:
LOCUS       BI698181               450 bp    mRNA    linear   EST 18-SEP-2001
DEFINITION  949031A01.y4 949 - Juvenile leaf and shoot cDNA from Steve Moose
            Zea mays cDNA, mRNA sequence.
ACCESSION   BI698181
VERSION     BI698181.1  GI:15660810
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 949031  row: A  column: 01.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Zea mays"
                     /cultivar="W64A"
                     /db_xref="taxon:4577"
                     /clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
                     Moose"
                     /tissue_type="immature leaf primordium and vegetative
                     meristem"
                     /dev_stage="4 stages from 3-13 days after imbibing"
                     /lab_host="E. coli XLOLR"
                     /note="Organ: juvenile vegetative shoots; Vector:
                     pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
                     of total RNA by weight from 4 tissue sources (see below)
                     were pooled, polyA+ RNA isolated, and cDNA synthesized for
                     EcoRI (5') and XhoI (3') directional cloning into lambda
                     Hybrizap vector from Stratagene. Tissue Sources: 1. Whole
                     shoots 3 days after sowing/imbibing in wet soil. 2. Basal
                     1.5 cm shoots 6 days after sowing - includes yellow
                     portions of developing leaves 1-5, primordia from 6-8, and
                     the vegetative apex. 3. Non-green portions of developing
                     leaves 4-5 and the vegetative apex,  including adult leaf
                     primordia, 9 days after sowing. 4. Partially expanded and
                     greening leaves 4-5 at 13 days after sowing."
BASE COUNT      111 a    108 c    112 g    119 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BI698181   ..
```

```
 Align seg 1/1  to: BI698181  from: 1  to: 450

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     128 AAGGCAACCGTCAAATTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 174

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     175 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 224

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     225 AGATG..................................................... 229

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   ::::::|||
     230 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 259

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                               ::::::||||||::::::      |||
     260 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 287

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::           ||||||||||||
     288 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 319

seq_name: gb_est1:AW018175

seq_documentation_block:
LOCUS       AW018175                459 bp    mRNA    linear   EST 13-SEP-1999
DEFINITION  614068C08.x1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW018175
VERSION     AW018175.1  GI:5871704
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614068  row: C  column: 08.
FEATURES             Location/Qualifiers
     source          1. .459
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      108 a    119 c    109 g    122 t      1 others
ORIGIN


alignment_scores:
            Quality:   75.50            Length:     94
              Ratio:  1.606              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532
```

```
alignment_block:
 US-10-048-197-2 x AW018175/rev   ..

 Align seg 1/1  to reverse of: AW018175  from: 1  to: 459

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     259 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 213

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     212 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 163

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     162 AGATG..................................................... 158

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   ::::::|||
     157 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 128

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                               ::::::||||||::::::      |||
     127 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 100

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::           ||||||||||||
      99 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 68

seq_name: gb_est1:AI834558

seq_documentation_block:
LOCUS       AI834558                460 bp    mRNA    linear   EST 02-FEB-2000
DEFINITION  606069H03.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI834558
VERSION     AI834558.1  GI:5468767
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606069  row: H  column: 03.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XLOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT      119 a    112 c     97 g    132 t
ORIGIN
```

```
alignment_scores:
            Quality:   75.50              Length:     94
              Ratio:  1.606                Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AI834558/rev   ..

 Align seg 1/1  to reverse of: AI834558  from: 1  to: 460

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     397 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 351

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |:::::|||||:::|||||:::|||   |||   ||||||:::|||:
     350 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 301

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     300 AGATG..................................................... 296

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||:::::::   :::::::|||
     295 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 266

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                               ::::::||||||:::::      |||
     265 .....................AAGAACTGGCAGAATGTGAGATGCCTTT 238

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::            ||||||||||||
     237 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 206

seq_name: gb_est2:BI245316

seq_documentation_block:
LOCUS       BI245316                 467 bp    mRNA    linear   EST 13-JUL-2001
DEFINITION  949031A01.x2 949 - Juvenile leaf and shoot cDNA from Steve Moose
            Zea mays cDNA, mRNA sequence.
ACCESSION   BI245316
VERSION     BI245316.1  GI:14717666
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 949031  row: A  column: 01.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Zea mays"
                     /cultivar="W64A"
                     /db_xref="taxon:4577"
                     /clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
                     Moose"
                     /tissue_type="immature leaf primordium and vegetative
                     meristem"
                     /dev_stage="4 stages from 3-13 days after imbibing"
                     /lab_host="E. coli XLOLR"
                     /note="Organ: juvenile vegetative shoots; Vector:
                     pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
                     of total RNA by weight from 4 tissue sources (see below)
                     were pooled, polyA+ RNA isolated, and cDNA synthesized for
                     EcoRI (5') and XhoI (3') directional cloning into lambda
                     Hybrizap vector from Stratagene. Tissue Sources: 1. Whole
                     shoots 3 days after sowing/imbibing in wet soil. 2. Basal
                     1.5 cm shoots 6 days after sowing - includes yellow
                     portions of developing leaves 1-5, primordia from 6-8, and
                     the vegetative apex. 3. Non-green portions of developing
                     leaves 4-5 and the vegetative apex,  including adult leaf
                     primordia, 9 days after sowing. 4. Partially expanded and
                     greening leaves 4-5 at 13 days after sowing."
BASE COUNT      116 a    111 c    116 g    124 t
ORIGIN


alignment_scores:
            Quality:   75.50              Length:     94
              Ratio:  1.606                Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BI245316   ..

 Align seg 1/1  to: BI245316  from: 1  to: 467

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     129 AAGGCAACCGTCAAATTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 175

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |:::::|||||:::|||||:::|||   |||   ||||||:::|||:
     176 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 225

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     226 AGATG..................................................... 230

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||:::::::   :::::::|||
     231 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 260

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                               ::::::||||||:::::      |||
     261 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 288

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::            ||||||||||||
     289 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 320

seq_name: gb_est2:BE511166

seq_documentation_block:
LOCUS       BE511166                 488 bp    mRNA    linear   EST 07-AUG-2000
DEFINITION  946058D02.x1 946 - tassel primordium prepared by Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   BE511166
VERSION     BE511166.1  GI:9732414
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
```

```
           855 California Ave, Palo Alto, CA 94304, USA
           Tel: 650 723 2227
           Fax: 650 725 8221
           Email: walbot@stanford.edu
           Plate: 946058  row: D  column: 02.
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Zea mays"
                     /cultivar="OH43"
                     /db_xref="taxon:4577"
                     /clone_lib="946 - tassel primordium prepared by Schmidt
                     lab"
                     /tissue_type="tassels"
                     /dev_stage="just after the transition from vegetative to
                     inflorescence development"
                     /lab_host="XLOLR"
                     /note="Organ: tassels; Vector: HybriZAP; Site_1: EcoRI;
                     Site_2: XhoI; George Chuck dissected immature tassels
                     between 1mm and 3mm.  Sharon  Stanfield prepared the cDNA
                     library in HybriZAP.  Sample insert size  range was 350 bp
                     to 3 Kb with a 1 Kb average."
BASE COUNT      129 a    121 c    116 g    122 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BE511166/rev   ..

 Align seg 1/1  to reverse of: BE511166  from: 1  to: 488

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     346 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 300

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     299 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 250

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     249 AGATG..................................................... 245

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                     |||:::::::|||::::::   ::::::|||
     244 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 215

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                       :     ::::::||||||::::::      |||
     214 ......................AAGAACTGGCAGAATGTGAGATGCCTCT 187

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::           ||||||||||||
     186 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 155

seq_name: gb_est2:BI431074

seq_documentation_block:
LOCUS       BI431074               491 bp    mRNA    linear   EST 20-AUG-2001
DEFINITION  949069B08.x1 949 - Juvenile leaf and shoot cDNA from Steve Moose
            Zea mays cDNA, mRNA sequence.
ACCESSION   BI431074
VERSION     BI431074.1  GI:15214850
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 949069  row: B  column: 08.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Zea mays"
                     /cultivar="W64A"
                     /db_xref="taxon:4577"
                     /clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
                     Moose"
                     /tissue_type="immature leaf primordium and vegetative
                     meristem"
                     /dev_stage="4 stages from 3-13 days after imbibing"
                     /lab_host="E. coli XLOLR"
                     /note="Organ: juvenile vegetative shoots; Vector:
                     pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
                     of total RNA by weight from 4 tissue sources (see below)
                     were pooled, polyA+ RNA isolated, and cDNA synthesized for
                     EcoRI (5') and XhoI (3') directional cloning into lambda
                     Hybrizap vector from Stratagene. Tissue Sources: 1. Whole
                     shoots 3 days after sowing/imbibing in wet soil. 2. Basal
                     1.5 cm shoots 6 days after sowing - includes yellow
                     portions of developing leaves 1-5, primordia from 6-8, and
                     the vegetative apex. 3. Non-green portions of developing
                     leaves 4-5 and the vegetative apex,  including adult leaf
                     primordia, 9 days after sowing. 4. Partially expanded and
                     greening leaves 4-5 at 13 days after sowing."
BASE COUNT      121 a    124 c    116 g    130 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BI431074/rev   ..

 Align seg 1/1  to reverse of: BI431074  from: 1  to: 491

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     303 AAGGCAACCGTCAAATTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 257

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     256 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 207

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     206 AGATG..................................................... 202

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                     |||:::::::|||::::::   ::::::|||
     201 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 172

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                       ::::::||||||::::::      |||
     171 ......................AAGAACTGGCAGAATGTGAGATGCCTCT 144

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::           ||||||||||||
```

```
    143 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 112

seq_name: gb_est1:AW037170

seq_documentation_block:
LOCUS       AW037170               492 bp    mRNA    linear   EST 15-SEP-1999
DEFINITION  614018H02.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW037170
VERSION     AW037170.1  GI:5895924
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614018  row: H  column: 02.
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      124 a    116 c    125 g    126 t      1 others
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AW037170   ..

 Align seg 1/1  to: AW037170  from: 1  to: 492

      7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
        :::   |||::::::      |||      :::|||      |||   ||
    154 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 200

     23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
        |::::::||||||:::||||||:::|||   |||   ||||||:::||||:
    201 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 250

     39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
        :::::
    251 AGATG..................................................... 255

     56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   ::::::|||
    256 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 285

     72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                              ::::::||||||::::::      |||

    286 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 313

     89 leIleSerPheLeuTrpValLysProTyrArg 99
        ::::::           ||||||||||||
    314 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 345

seq_name: gb_est2:BI478440

seq_documentation_block:
LOCUS       BI478440               512 bp    mRNA    linear   EST 27-AUG-2001
DEFINITION  949069B08.y1 949 - Juvenile leaf and shoot cDNA from Steve Moose
            Zea mays cDNA, mRNA sequence.
ACCESSION   BI478440
VERSION     BI478440.1  GI:15312868
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 949069  row: B  column: 08.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Zea mays"
                     /cultivar="W64A"
                     /db_xref="taxon:4577"
                     /clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
                     Moose"
                     /tissue_type="immature leaf primordium and vegetative
                     meristem"
                     /dev_stage="4 stages from 3-13 days after imbibing"
                     /lab_host="E. coli XLOLR"
                     /note="Organ: juvenile vegetative shoots; Vector:
                     pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
                     of total RNA by weight from 4 tissue sources (see below)
                     were pooled, polyA+ RNA isolated, and cDNA synthesized for
                     EcoRI (5') and XhoI (3') directional cloning into lambda
                     Hybrizap vector from Stratagene. Tissue Sources: 1. Whole
                     shoots 3 days after sowing/imbibing in wet soil. 2. Basal
                     1.5 cm shoots 6 days after sowing - includes yellow
                     portions of developing leaves 1-5, primordia from 6-8, and
                     the vegetative apex. 3. Non-green portions of developing
                     leaves 4-5 and the vegetative apex,  including adult leaf
                     primordia, 9 days after sowing. 4. Partially expanded and
                     greening leaves 4-5 at 13 days after sowing."
BASE COUNT      139 a    122 c    126 g    125 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BI478440   ..

 Align seg 1/1  to: BI478440  from: 1  to: 512

      7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
        :::   |||::::::      |||      :::|||      |||   ||
```

```
     148 AAGGCAACCGTCAAATTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 194

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |:::::|||||||:::|||||||:::|||   |||   |||||||:::|||:
     195 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 244

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     245 AGATG..................................................... 249

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                       |||:::::::|||::::::   ::::::|||
     250 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 279

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                               ::::::||||||::::::      |||
     280 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 307

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::            ||||||||||||
     308 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 339
```

seq_name: gb_est1:AW171836

seq_documentation_block:

```
LOCUS       AW171836                513 bp    mRNA    linear   EST 12-NOV-1999
DEFINITION  618051B11.x1 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA
            sequence.
ACCESSION   AW171836
VERSION     AW171836.1  GI:6403361
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 513)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 618051  row: B  column: 11.
FEATURES             Location/Qualifiers
     source          1. .513
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="618 - Inbred Tassel cDNA Library"
                     /tissue_type="tassel"
                     /dev_stage="tassel length from 0.1 to 2.5 cm"
                     /lab_host="XLOLR"
                     /note="Organ: tassel; Vector: pAD-GAL4-2.1 (Hybrizap);
                     Inbred tassel library from Schmidt lab"
BASE COUNT      114 a    137 c    128 g    134 t
ORIGIN
```

alignment_scores:

```
            Quality:   75.50            Length:     94
              Ratio:  1.606              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532
```

alignment_block:

```
 US-10-048-197-2 x AW171836/rev   ..

 Align seg 1/1  to reverse of: AW171836  from: 1  to: 513
```

```
       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     240 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 194

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::|||||||:::|||||||:::|||   |||   |||||||:::|||:
     193 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 144

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     143 AGATG..................................................... 139

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                       |||:::::::|||::::::   ::::::|||
     138 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 109

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                               ::::::||||||::::::      |||
     108 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 81

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::            ||||||||||||
      80 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 49
```

seq_name: gb_est1:AI586595

seq_documentation_block:

```
LOCUS       AI586595                514 bp    mRNA    linear   EST 07-APR-1999
DEFINITION  486049C06.x4 486 - leaf primordia cDNA library from Hake lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI586595
VERSION     AI586595.1  GI:4572946
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 514)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 486049  row: C  column: 06.
FEATURES             Location/Qualifiers
     source          1. .514
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone_lib="486 - leaf primordia cDNA library from Hake
                     lab"
                     /tissue_type="leaf primordia"
                     /dev_stage="P7-P11 leaf"
                     /lab_host="E.coli XL1-Blue MFR'"
                     /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
                     library."
BASE COUNT      138 a    123 c    113 g    140 t
ORIGIN
```

alignment_scores:

```
            Quality:   75.50            Length:     94
              Ratio:  1.606              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532
```

```
alignment_block:
 US-10-048-197-2 x AI586595/rev   ..

 Align seg 1/1  to reverse of: AI586595  from: 1  to: 514

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||       :::|||       |||   ||
     376 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTTCTTTGCATGGGT...GT 330

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     329 TGCCGTTGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 280

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     279 AGATG..................................................... 275

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   :::::::|||
     274 ..............AGCGTCAACTTCCTTGTTTCCCTTCTGAAG..... 245

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                              ::::::||||||::::::      |||
     244 ......................AAGAACTGGCAGAATGTGAGATGCCTCT 217

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::            ||||||||||||
     216 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 185

seq_name: gb_est2:BI643532

seq_documentation_block:
LOCUS       BI643532               518 bp    mRNA    linear   EST 10-SEP-2001
DEFINITION  949079C10.x1 949 - Juvenile leaf and shoot cDNA from Steve Moose
            Zea mays cDNA, mRNA sequence.
ACCESSION   BI643532
VERSION     BI643532.1  GI:15545738
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 518)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 949079  row: C  column: 10.
FEATURES             Location/Qualifiers
     source          1. .518
                     /organism="Zea mays"
                     /cultivar="W64A"
                     /db_xref="taxon:4577"
                     /clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
                     Moose"
                     /tissue_type="immature leaf primordium and vegetative
                     meristem"
                     /dev_stage="4 stages from 3-13 days after imbibing"
                     /lab_host="E. coli XLOLR"
                     /note="Organ: juvenile vegetative shoots; Vector:
                     pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
                     of total RNA by weight from 4 tissue sources (see below)
                     were pooled, polyA+ RNA isolated, and cDNA synthesized for
                     EcoRI (5') and XhoI (3') directional cloning into lambda
                     Hybrizap vector from Stratagene. Tissue Sources: 1. Whole
                     shoots 3 days after sowing/imbibing in wet soil. 2. Basal
                     1.5 cm shoots 6 days after sowing - includes yellow
                     portions of developing leaves 1-5, primordia from 6-8, and
                     the vegetative apex. 3. Non-green portions of developing
                     leaves 4-5 and the vegetative apex,  including adult leaf
                     primordia, 9 days after sowing. 4. Partially expanded and
                     greening leaves 4-5 at 13 days after sowing."
BASE COUNT      138 a    130 c    122 g    128 t
ORIGIN


alignment_scores:
           Quality:   75.50            Length:     94
             Ratio:  1.606              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BI643532/rev   ..

 Align seg 1/1  to reverse of: BI643532  from: 1  to: 518

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||       :::|||       |||   ||
     358 AAGGCAACCGTCAAATTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 312

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     311 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 262

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     261 AGATG..................................................... 257

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   :::::::|||
     256 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 227

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                              ::::::||||||::::::      |||
     226 ......................AAGAACTGGCAGAATGTGAGATGCCTCT 199

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::            ||||||||||||
     198 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 167

seq_name: gb_est1:AI629856

seq_documentation_block:
LOCUS       AI629856               529 bp    mRNA    linear   EST 26-APR-1999
DEFINITION  486041B10.x2 486 - leaf primordia cDNA library from Hake lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI629856
VERSION     AI629856.1  GI:4681186
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 486041  row: B  column: 10.
FEATURES             Location/Qualifiers
```

```
     source          1. .529
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone_lib="486 - leaf primordia cDNA library from Hake
                     lab"
                     /tissue_type="leaf primordia"
                     /dev_stage="P7-P11 leaf"
                     /lab_host="E.coli XL1-Blue MFR'"
                     /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
                     library."
BASE COUNT      142 a    126 c    114 g    147 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AI629856/rev   ..

 Align seg 1/1  to reverse of: AI629856  from: 1  to: 529

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     397 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTTCTTTGCATGGGT...GT 351

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     350 TGCCGTTGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 301

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     300 AGATG..................................................... 296

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                       |||:::::::|||::::::   ::::::|||
     295 ..............AGCGTCAACTTCCTTGTTTCCCTTCTGAAG..... 266

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                               ::::::||||||::::::      |||
     265 ......................AAGAACTGGCAGAATGTGAGATGCCTCT 238

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::            ||||||||||||
     237 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 206

seq_name: gb_est2:BI180391

seq_documentation_block:
LOCUS       BI180391                533 bp    mRNA    linear   EST 09-JUL-2001
DEFINITION  949031A01.x1 949 - Juvenile leaf and shoot cDNA from Steve Moose
            Zea mays cDNA, mRNA sequence.
ACCESSION   BI180391
VERSION     BI180391.1  GI:14646202
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 533)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 949031  row: A  column: 01.
FEATURES             Location/Qualifiers
     source          1. .533
                     /organism="Zea mays"
                     /cultivar="W64A"
                     /db_xref="taxon:4577"
                     /clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
                     Moose"
                     /tissue_type="immature leaf primordium and vegetative
                     meristem"
                     /dev_stage="4 stages from 3-13 days after imbibing"
                     /lab_host="E. coli XLOLR"
                     /note="Organ: juvenile vegetative shoots; Vector:
                     pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
                     of total RNA by weight from 4 tissue sources (see below)
                     were pooled, polyA+ RNA isolated, and cDNA synthesized for
                     EcoRI (5') and XhoI (3') directional cloning into lambda
                     Hybrizap vector from Stratagene. Tissue Sources: 1. Whole
                     shoots 3 days after sowing/imbibing in wet soil. 2. Basal
                     1.5 cm shoots 6 days after sowing - includes yellow
                     portions of developing leaves 1-5, primordia from 6-8, and
                     the vegetative apex. 3. Non-green portions of developing
                     leaves 4-5 and the vegetative apex,  including adult leaf
                     primordia, 9 days after sowing. 4. Partially expanded and
                     greening leaves 4-5 at 13 days after sowing."
BASE COUNT      142 a    133 c    124 g    134 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BI180391/rev   ..

 Align seg 1/1  to reverse of: BI180391  from: 1  to: 533

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     364 AAGGCAACCGTCAAATTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 318

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     317 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 268

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     267 AGATG..................................................... 263

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                       |||:::::::|||::::::   ::::::|||
     262 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 233

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                               ::::::||||||::::::      |||
     232 ......................AAGAACTGGCAGAATGTGAGATGCCTCT 205

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::            ||||||||||||
     204 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 173

seq_name: gb_est1:AW090901

seq_documentation_block:
LOCUS       AW090901                554 bp    mRNA    linear   EST 18-OCT-1999
DEFINITION  614068C08.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW090901
VERSION     AW090901.1  GI:6056511
```

```
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 554)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614068  row: C  column: 08.
FEATURES             Location/Qualifiers
     source          1. .554
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      143 a    129 c    142 g    140 t
ORIGIN


alignment_scores:
            Quality:   75.50            Length:     94
              Ratio:  1.606              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AW090901   ..

 Align seg 1/1  to: AW090901  from: 1  to: 554

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     218 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 264

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     265 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 314

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     315 AGATG..................................................... 319

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   :::::::|||
     320 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 349

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                              ::::::||||||::::::      |||
     350 ......................AAGAACTGGCAGAATGTGAGATGCCTCT 377

      89 leIleSerPheLeuTrpValLysProTyrArg 99
         ::::::          ||||||||||||
     378 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 409

seq_name: gb_est2:BI675165

seq_documentation_block:
LOCUS       BI675165                 557 bp    mRNA    linear   EST 12-SEP-2001
```

```
DEFINITION  949079C10.x2 949 - Juvenile leaf and shoot cDNA from Steve Moose
            Zea mays cDNA, mRNA sequence.
ACCESSION   BI675165
VERSION     BI675165.1  GI:15590549
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 557)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 949079  row: C  column: 10.
FEATURES             Location/Qualifiers
     source          1. .557
                     /organism="Zea mays"
                     /cultivar="W64A"
                     /db_xref="taxon:4577"
                     /clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
                     Moose"
                     /tissue_type="immature leaf primordium and vegetative
                     meristem"
                     /dev_stage="4 stages from 3-13 days after imbibing"
                     /lab_host="E. coli XLOLR"
                     /note="Organ: juvenile vegetative shoots; Vector:
                     pAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
                     of total RNA by weight from 4 tissue sources (see below)
                     were pooled, polyA+ RNA isolated, and cDNA synthesized for
                     EcoRI (5') and XhoI (3') directional cloning into lambda
                     Hybrizap vector from Stratagene. Tissue Sources: 1. Whole
                     shoots 3 days after sowing/imbibing in wet soil. 2. Basal
                     1.5 cm shoots 6 days after sowing - includes yellow
                     portions of developing leaves 1-5, primordia from 6-8, and
                     the vegetative apex. 3. Non-green portions of developing
                     leaves 4-5 and the vegetative apex,  including adult leaf
                     primordia, 9 days after sowing. 4. Partially expanded and
                     greening leaves 4-5 at 13 days after sowing."
BASE COUNT      143 a    143 c    124 g    147 t
ORIGIN


alignment_scores:
            Quality:   75.50            Length:     94
              Ratio:  1.606              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BI675165/rev   ..

 Align seg 1/1  to reverse of: BI675165  from: 1  to: 557

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     337 AAGGCAACCGTCAAATTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 291

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     290 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 241

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     240 AGATG..................................................... 236
```

```
     56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                     |||::::::|||::::::   :::::::|||
    235 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 206

     72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                           ::::::||||||::::::      |||
    205 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 178

     89 leIleSerPheLeuTrpValLysProTyrArg 99
        ::::::            ||||||||||||
    177 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 146

seq_name: gb_est1:AI714378

seq_documentation_block:
LOCUS       AI714378               559 bp    mRNA    linear   EST 02-FEB-2000
DEFINITION  606013A04.x2 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI714378
VERSION     AI714378.1  GI:5018185
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606013  row: A  column: 04.
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XLOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT      144 a    136 c    126 g    153 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AI714378/rev   ..

 Align seg 1/1  to reverse of: AI714378  from: 1  to: 559

      7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
        :::   |||::::::      |||      :::|||      |||   ||
    393 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 347

     23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
        |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
    346 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 297

     39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
        :::::
    296 AGATG..................................................... 292

     56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                     |||::::::|||::::::   ::::::|||
    291 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 262

     72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                           ::::::||||||::::::      |||
    261 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 234

     89 leIleSerPheLeuTrpValLysProTyrArg 99
        ::::::            ||||||||||||
    233 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 202

seq_name: gb_est1:AW566286

seq_documentation_block:
LOCUS       AW566286               563 bp    mRNA    linear   EST 10-MAR-2000
DEFINITION  660069A03.y1 660 - Mixed stages of anther and pollen Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW566286
VERSION     AW566286.1  GI:7227645
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 660069  row: A  column: 03.
FEATURES             Location/Qualifiers
     source          1. .563
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="660 - Mixed stages of anther and pollen"
                     /tissue_type="whole premieotic anthers to pollen shed"
                     /dev_stage="premieotic anthers to pollen shed"
                     /lab_host="XLOLR"
                     /note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
                     Site_2: XhoI; Anther and pollen cDNA library.
                     Directionally sequenced with 5' end at the EcoRI site.
                     Created by Amie Franklin."
BASE COUNT      146 a    136 c    153 g    128 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AW566286   ..

 Align seg 1/1  to: AW566286  from: 1  to: 563

      7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
        :::   |||::::::      |||      :::|||      |||   ||
```

```
     324 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 370

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |:::::|||||:::|||||:::|||   |||   ||||||:::|||:
     371 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 420

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     421 AGATG..................................................... 425

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                       |||:::::|||::::::   :::::::|||
     426 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 455

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                              ::::::||||||::::::      |||
     456 ......................AAGAACTGGCAGAATGTGAGATGCCTCT 483

      89 leIleSerPheLeuTrpValLysProTyrArg 99
         ::::::           ||||||||||||
     484 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 515

seq_name: gb_est2:BF728138

seq_documentation_block:
LOCUS       BF728138                587 bp    mRNA    linear   EST 09-JAN-2001
DEFINITION  1000056E09.x2 1000 - Unigene I from Maize Genome Project Zea mays
            cDNA, mRNA sequence.
ACCESSION   BF728138
VERSION     BF728138.1  GI:12045908
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 1000056  row: E  column: 09.
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="Zea mays"
                     /db_xref="dbEST:707024C08.x1"
                     /db_xref="taxon:4577"
                     /clone_lib="1000 - Unigene I from Maize Genome Project"
                     /note="This library represents the unique ESTs found in
                     the first round of EST sequencing at Stanford University
                     for the maize genome project. Sequences are present from
                     libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683
                     , 687, 707, and 945. Contigs were assembled using TIGR's
                     CAP program and a representative EST from each contig was
                     selected for the Unigene set. All singlets were also
                     selected."
BASE COUNT      157 a    144 c    128 g    156 t      2 others
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
```

```
US-10-048-197-2 x BF728138/rev   ..

Align seg 1/1  to reverse of: BF728138  from: 1  to: 587

      7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
        :::   |||::::::      |||      :::|||      |||   ||
    385 AAGGCAACAGTCAAGTTCCAACTCAAGAAGGTCCTTTGCATGGGT...GT 339

     23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
        |:::::|||||:::|||||:::|||   |||   ||||||:::|||:
    338 CGCTGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 289

     39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
        :::::
    288 AGATG..................................................... 284

     56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                      |||:::::|||::::::   :::::::|||
    283 ..............AGCGTCAACTTCCTTGTTTCCCTTCTGAAG..... 254

     72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                             ::::::||||||::::::      |||
    253 ......................AAGAACTGGCAGAATGTGAGATGCCTGT 226

     89 leIleSerPheLeuTrpValLysProTyrArg 99
        ::::::           ||||||||||||
    225 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 194

seq_name: gb_est2:BG842695

seq_documentation_block:
LOCUS       BG842695                599 bp    mRNA    linear   EST 29-MAY-2001
DEFINITION  MEST39-A03.T3 ISUM4-TN Zea mays cDNA clone MEST39-A03 3', mRNA
            sequence.
ACCESSION   BG842695
VERSION     BG842695.2  GI:14244757
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 599)
  AUTHORS   Qiu,F., Cui,F., Guo,L., Ashlock,D.A, Wen,T.J. and Schnable,P.S.
  TITLE     Expressed Sequence Tags from B73 Maize Seedlings and Silks
  JOURNAL   Unpublished (2001)
COMMENT     On May 25, 2001 this sequence version replaced gi:14209017.
            Contact: Patrick S. Schnable
            Schnable Laboratory
            Iowa State University
            G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
            Tel: 515-294-0975
            Fax: 515-294-2299
            Email: schnable@iastate.edu
            PCR PRimers
            FORWARD: T7-1 (AA TAC GAC TCA CTA TAG)
            BACKWARD: T3 (ATT AAC CCT CAC TAA AG)
            Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).
FEATURES             Location/Qualifiers
     source          1. .599
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="MEST39-A03"
                     /clone_lib="ISUM4-TN"
                     /tissue_type="Seedling and silk"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3PAC; Site_1: EcoRI; Site_2: NotI;
                     ds-cDNA molecules were generated as follows. First-strand
                     cDNA was prepared from oligo-dT selected mRNA by priming
                     with a NotI  oligo-dT primer (5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT). The
                     resulting DNA:RNA hybrid was treated with RNase H and used
```

```
                as a  template for DNA PolI-catalyzed second strand
                synthesis. After the  addition of EcoRI adaptors, the
                ds-cDNAs were digested with NotI and  size-selected. The
                resulting molecules were directionally cloned  into the
                EcoRI and NotI sites of the pT7T3PAC vector. The library
                then went through one round of normalization to CoT value
                of 5 based  on the methods of Marcelo Bento Soares (Genome
                Research 6: 791-806,  1996)."
BASE COUNT      160 a    141 c    129 g    169 t
ORIGIN


alignment_scores:
           Quality:   75.50            Length:     94
             Ratio:  1.606              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BG842695/rev   ..

 Align seg 1/1  to reverse of: BG842695  from: 1  to: 599

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::     |||      :::|||      |||   ||
     411 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTTCTTTGCATGGGT...GT 365

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     364 TGCCGTTGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 315

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     314 AGATG..................................................... 310

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   :::::::|||
     309 ..............AGCGTCAACTTCCTTGTTTCCCTTCTGAAG..... 280

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                           ::::::||||||::::::      |||
     279 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 252

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::            ||||||||||||
     251 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 220

seq_name: gb_est1:AI691931

seq_documentation_block:
LOCUS       AI691931               602 bp    mRNA    linear   EST 02-FEB-2000
DEFINITION  606007A02.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI691931
VERSION     AI691931.1  GI:4966852
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606007  row: A  column: 02.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XLOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT      151 a    144 c    133 g    174 t
ORIGIN


alignment_scores:
           Quality:   75.50            Length:     94
             Ratio:  1.606              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AI691931/rev   ..

 Align seg 1/1  to reverse of: AI691931  from: 1  to: 602

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::     |||      :::|||      |||   ||
     414 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 368

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::|||   |||   ||||||:::|||:
     367 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 318

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     317 AGATG..................................................... 313

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   :::::::|||
     312 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 283

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                           ::::::||||||::::::      |||
     282 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 255

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::            ||||||||||||
     254 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 223

seq_name: gb_est2:BM079588

seq_documentation_block:
LOCUS       BM079588               607 bp    mRNA    linear   EST 14-NOV-2001
DEFINITION  MEST96-F07.T3 ISUM4-TN Zea mays cDNA clone MEST96-F07 3', mRNA
            sequence.
ACCESSION   BM079588
VERSION     BM079588.1  GI:16926520
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 607)
  AUTHORS   Qiu,F., Cui,F., Guo,L., Ashlock,D.A, Wen,T.J. and Schnable,P.S.
  TITLE     Expressed Sequence Tags from B73 Maize Seedlings and Silks
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Patrick S. Schnable
            Schnable Laboratory
            Iowa State University
            G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA
```

```
           Tel: 515-294-0975
           Fax: 515-294-2299
           Email: schnable@iastate.edu
           Individual basecall and confidence value were assigned using the
           Phred software,
           (<http://depts.washington.edu/ventures/collabtr/direct/index.htm>#b
           rt). Overall sequence quality assessment and vector trimming were
           conducted using the Lucy software (<http://www.tigr.org/softlab/>).
           Lucy parameters were set to ensure an overall trimmed quality of
           97.5% or better without any vector fragments in the chosen
           high-quality region of each sequence. Low-quality bases between the
           poly-T and the high-quality region were replaced with N's to serve
           as spacers.
           PCR PRimers
           FORWARD: Forward PCR primer sequence, primer T7-1 (AA TAC GAC TCA
           CTA TAG)
           BACKWARD: Backward PCR primer sequence, primer T3 (ATT AAC CCT CAC
           TAA AG)
           Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).
FEATURES             Location/Qualifiers
     source          1. .607
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="MEST96-F07"
                     /clone_lib="ISUM4-TN"
                     /tissue_type="Seedling and silk"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3PAC; Site_1: EcoRI; Site_2: NotI;
                     ds-cDNA molecules were generated as follows. First-strand
                     cDNA was prepared from oligo-dT selected mRNA by priming
                     with a NotI  oligo-dT primer (5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTT). The
                     resulting DNA:RNA hybrid was treated with RNase H and used
                     as a  template for DNA PolI-catalyzed second strand
                     synthesis. After the  addition of EcoRI adaptors, the
                     ds-cDNAs were digested with NotI and  size-selected. The
                     resulting molecules were directionally cloned  into the
                     EcoRI and NotI sites of the pT7T3PAC vector. The library
                     then went through one round of normalization to CoT value
                     of 5 based  on the methods of Marcelo Bento Soares (Genome
                     Research 6: 791-806,  1996)."
BASE COUNT      158 a    144 c    128 g    177 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BM079588/rev   ..

 Align seg 1/1  to reverse of: BM079588  from: 1  to: 607

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     417 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTTCTTTGCATGGGT...GT 371

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     370 TGCCGTTGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 321

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     320 AGATG..................................................... 316

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   ::::::|||
     315 ..............AGCGTCAACTTCCTTGTTTCCCTTCTGAAG..... 286

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                                ::::::||||||::::::      |||
     285 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 258

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::            ||||||||||||
     257 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 226

seq_name: gb_est1:AW324655

seq_documentation_block:
LOCUS       AW324655                611 bp    mRNA    linear   EST 27-JAN-2000
DEFINITION  707036H01.x1 707 - Mixed adult tissues from Walbot lab (SK) Zea
            mays cDNA, mRNA sequence.
ACCESSION   AW324655
VERSION     AW324655.1  GI:6760556
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 707036  row: H  column: 01.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="707 - Mixed adult tissues from Walbot lab (SK
                     )"
                     /tissue_type="tassel, kernel, silk, husk, root, leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: tassel, kernel, silk, husk, root, leaf;
                     Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
                     differentiated maize tissues from an active  Mutator
                     plant. Tissue ratio is 4/2/1/1/1/1 (tassel, kernel, silk,
                     husk, root, leaf).  Unidirectionally cloned."
BASE COUNT      155 a    153 c    133 g    170 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AW324655/rev   ..

 Align seg 1/1  to reverse of: AW324655  from: 1  to: 611

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     388 AAGGCAACAGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGT...GT 342

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     341 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 292

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
```

```
     291 AGATG..................................................... 287

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||:::::|||::::::   :::::::|||
     286 ..............AGCGTCAACTTCCTTGTTTCCCTTCTGAAG..... 257

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                              ::::::||||||::::::      |||
     256 ......................AAGAACTGGCAGAATGTGAGATGCCTGT 229

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::           ||||||||||||
     228 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 197

seq_name: gb_est1:AI711637

seq_documentation_block:
LOCUS       AI711637               613 bp    mRNA    linear   EST 02-FEB-2000
DEFINITION  605058E06.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
            cDNA, mRNA sequence.
ACCESSION   AI711637
VERSION     AI711637.1  GI:5005575
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 605058  row: E  column: 06.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
                     /tissue_type="nucellar, embryo, and endosperm"
                     /dev_stage="10-14 days post-pollination"
                     /lab_host="DH5(alpha)"
                     /note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI;
                     Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
                     lab"
BASE COUNT      158 a    158 c    138 g    159 t
ORIGIN


alignment_scores:
            Quality:   75.50            Length:     94
              Ratio:  1.606              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AI711637/rev   ..

 Align seg 1/1  to reverse of: AI711637  from: 1  to: 613

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||  ||
     362 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 316

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
```

```
     315 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 266

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     265 AGATG..................................................... 261

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||:::::|||::::::   :::::::|||
     260 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 231

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                              ::::::||||||::::::      |||
     230 ......................AAGAACTGGCAGAATGTGAGATGCCTCT 203

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::           ||||||||||||
     202 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 171

seq_name: gb_est1:AW927882

seq_documentation_block:
LOCUS       AW927882               614 bp    mRNA    linear   EST 30-MAY-2000
DEFINITION  945010D06.Y1 945 - Mixed adult tissues from Walbot lab, same as 707
            (SK) Zea mays cDNA, mRNA sequence.
ACCESSION   AW927882
VERSION     AW927882.1  GI:8103229
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 945010  row: D  column: 06.
FEATURES             Location/Qualifiers
     source          1. .614
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="945 - Mixed adult tissues from Walbot lab,
                     same as 707 (SK)"
                     /tissue_type="tassel, kernal, silk, husk, root, leaf"
                     /dev_stage="fully-grown"
                     /lab_host="DH10B"
                     /note="Organ: tassel, kernal, silk, husk, root, leaf;
                     Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
                     differentiated maize tissues from an active Mutator plant.
                     Tissue ratio is 4:2:1:1:1:1 (tassel, kernel, silk, husk,
                     root, leaf). Unidirectionally cloned.  New library number
                     given to library 707  for additional sequencing."
BASE COUNT      174 a    133 c    149 g    158 t
ORIGIN


alignment_scores:
            Quality:   75.50            Length:     94
              Ratio:  1.606              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AW927882   ..
```

```
Align seg 1/1  to: AW927882  from: 1  to: 614

      7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
        :::   |||::::::      |||      :::|||      |||   ||
    200 AAGGCAACAGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGT...GT 246

     23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
        |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
    247 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 296

     39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
        :::::
    297 AGATG..................................................... 301

     56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                   |||::::::|||::::::   ::::::|||
    302 ..............AGCGTCAACTTCCTTGTTTCCCTTCTGAAG..... 331

     72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                             ::::::||||||:::::      |||
    332 .....................AAGAACTGGCAGAATGTGAGATGCCTGT 359

     89 leIleSerPheLeuTrpValLysProTyrArg 99
          ::::::           ||||||||||||
    360 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 391

seq_name: gb_est1:AI820102

seq_documentation_block:
LOCUS       AI820102               624 bp    mRNA    linear   EST 02-FEB-2000
DEFINITION  605086A01.x3 605 - Endosperm cDNA library from Schmidt lab Zea mays
            cDNA, mRNA sequence.
ACCESSION   AI820102
VERSION     AI820102.1  GI:5439181
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 605086  row: A  column: 01.
FEATURES             Location/Qualifiers
     source          1. .624
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
                     /tissue_type="nucellar, embryo, and endosperm"
                     /dev_stage="10-14 days post-pollination"
                     /lab_host="DH5(alpha)"
                     /note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI;
                     Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
                     lab"
BASE COUNT      161 a    162 c    139 g    160 t      2 others
ORIGIN


alignment_scores:
           Quality:   75.50            Length:     94
             Ratio:  1.606              Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532
```

```
alignment_block:
 US-10-048-197-2 x AI820102/rev   ..

 Align seg 1/1  to reverse of: AI820102  from: 1  to: 624

      7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
        :::   |||::::::      |||      :::|||      |||   ||
    361 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 315

     23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
        |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
    314 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 265

     39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
        :::::
    264 AGATG..................................................... 260

     56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                   |||::::::|||::::::   ::::::|||
    259 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 230

     72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                             ::::::||||||:::::      |||
    229 .....................AAGAACTGGCAGAATGTGAGATGCCTCT 202

     89 leIleSerPheLeuTrpValLysProTyrArg 99
          ::::::           ||||||||||||
    201 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 170

seq_name: gb_est2:BE509707

seq_documentation_block:
LOCUS       BE509707               634 bp    mRNA    linear   EST 07-AUG-2000
DEFINITION  946020H06.x1 946 - tassel primordium prepared by Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   BE509707
VERSION     BE509707.1  GI:9730955
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 634)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 946020  row: H  column: 06.
FEATURES             Location/Qualifiers
     source          1. .634
                     /organism="Zea mays"
                     /cultivar="OH43"
                     /db_xref="taxon:4577"
                     /clone_lib="946 - tassel primordium prepared by Schmidt
                     lab"
                     /tissue_type="tassels"
                     /dev_stage="just after the transition from vegetative to
                     inflorescence development"
                     /lab_host="XLOLR"
                     /note="Organ: tassels; Vector: HybriZAP; Site_1: EcoRI;
                     Site_2: XhoI; George Chuck dissected immature tassels
                     between 1mm and 3mm.  Sharon  Stanfield prepared the cDNA
                     library in HybriZAP.  Sample insert size  range was 350 bp
                     to 3 Kb with a 1 Kb average."
```

```
BASE COUNT      164 a    167 c    139 g    164 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x BE509707/rev   ..

 Align seg 1/1  to reverse of: BE509707  from: 1  to: 634

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     366 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 320

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     319 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 270

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     269 AGATG..................................................... 265

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   ::::::|||
     264 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 235

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                              ::::::||||||::::::      |||
     234 ......................AAGAACTGGCAGAATGTGAGATGCCTCT 207

      89 leIleSerPheLeuTrpValLysProTyrArg 99
         ::::::             ||||||||||||
     206 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 175

seq_name: gb_est1:AI665346

seq_documentation_block:
LOCUS       AI665346               638 bp    mRNA    linear   EST 02-FEB-2000
DEFINITION  605010B03.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
            cDNA, mRNA sequence.
ACCESSION   AI665346
VERSION     AI665346.2  GI:4804476
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 638)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     On May 11, 1999 this sequence version replaced gi:4776343.
            Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 605010  row: B  column: 03.
FEATURES             Location/Qualifiers
     source          1. .638
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
                     /tissue_type="nucellar, embryo, and endosperm"
                     /dev_stage="10-14 days post-pollination"
                     /lab_host="DH5(alpha)"
                     /note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI;
                     Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
                     lab"
BASE COUNT      165 a    167 c    140 g    166 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AI665346/rev   ..

 Align seg 1/1  to reverse of: AI665346  from: 1  to: 638

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     373 AAGGCAACCGTCAAGTTCCAGCTCAAGAAGGTCCTTTGCATGGGC...GT 327

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     326 CGCCGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 277

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     276 AGATG..................................................... 272

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   ::::::|||
     271 ..............AGCGTGAACTTCCTTGTTTCCCTTCTGAAG..... 242

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                              ::::::||||||::::::      |||
     241 ......................AAGAACTGGCAGAATGTGAGATGCCTCT 214

      89 leIleSerPheLeuTrpValLysProTyrArg 99
         ::::::             ||||||||||||
     213 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 182

seq_name: gb_est1:AW313326

seq_documentation_block:
LOCUS       AW313326               650 bp    mRNA    linear   EST 24-JAN-2000
DEFINITION  707024C08.x1 707 - Mixed adult tissues from Walbot lab (SK) Zea
            mays cDNA, mRNA sequence.
ACCESSION   AW313326
VERSION     AW313326.1  GI:6742511
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 650)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 707024  row: C  column: 08.
FEATURES             Location/Qualifiers
     source          1. .650
                     /organism="Zea mays"
```

```
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="707 - Mixed adult tissues from Walbot lab (SK
                     )"
                     /tissue_type="tassel, kernel, silk, husk, root, leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: tassel, kernel, silk, husk, root, leaf;
                     Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
                     differentiated maize tissues from an active  Mutator
                     plant. Tissue ratio is 4/2/1/1/1/1 (tassel, kernel, silk,
                     husk, root, leaf).  Unidirectionally cloned."
BASE COUNT      166 a    160 c    139 g    185 t
ORIGIN


alignment_scores:
            Quality:   75.50             Length:     94
              Ratio:  1.606               Gaps:      4
 Percent Similarity:  50.000   Percent Identity: 25.532

alignment_block:
 US-10-048-197-2 x AW313326/rev   ..

 Align seg 1/1  to reverse of: AW313326  from: 1  to: 650

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         :::   |||::::::      |||      :::|||      |||   ||
     425 AAGGCAACAGTCAAGTTCCAACTCAAGAAGGTCCTTTGCATGGGT...GT 379

      23 lSerThrGlyAsnValAlaMetLysGlu...GlnAsnGlnGlnThrIleG 39
         |::::::||||||:::||||||:::|||   |||   ||||||:::|||:
     378 CGCTGTCGGCAACTTGGCGATGGAGGAGAAGCAGATCCAGCAGAACATCC 329

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg 55
         :::::
     328 AGATG..................................................... 324

      56 PheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGl 72
                    |||::::::|||::::::   ::::::|||
     323 ..............AGCGTCAACTTCCTTGTTTCCCTTCTGAAG..... 294

      72 yHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuI 89
                             ::::::||||||::::::      |||
     293 ......................AAGAACTGGCAGAATGTGAGATGCCTGT 266

      89 leIleSerPheLeuTrpValLysProTyrArg 99
           ::::::            ||||||||||||
     265 ACGTCAAGAGCACCATGGGGAAGCCATACAGG 234
```

```
OM of: US-10-048-197-2 to: Issued_Patents_NA:*   out_format : pfs

Date: Sep 18, 2002  5:15 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US10048197/runat_17092002_135850_22630/app_query.fasta_1.168
 -DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=std.rni
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL
 -OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US10048197_@CGN1_1_74 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-10-048-197-2
Query length: 111
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 79.290000



score_list:
Sequence           Strd Orig    ZScore   EScore Len  ! Documentation  ..
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-332-576-1 +   67.50  135.99  6.92      2100  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-13672-1 +   67.50  135.99  6.92      2100
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-023-082A-23 -   67.50  128.04  19.18     4177  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-988-856B-3 -   66.50  135.83  7.07      1700  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-01806-3 -   66.50  135.83  7.07      1700
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-196-520-5 +   66.00  136.75  6.28      1402  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US96-05320A-333 +   65.50  149.88  1.16      40
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-371-377-16 +   64.50  128.02  19.23     2128  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-895-300A-2 +   63.50  133.70  9.28      1039  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-367-2 +   63.50  133.70  9.28      1039  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-05136-2 +   63.50  133.70  9.28      1039
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-10519-1 +   63.50  133.70  9.28      1039
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-013-810-1 +   63.00  136.22  6.72      746  !  S
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-058-489-34 +   63.00  106.13  318.86    10091  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-366-783-3 +   61.50  132.03  11.51     765  !  S
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-514-302-1 +   61.50  107.95  252.18    6142  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-309-487-25 -   61.00  116.80  81.14     2553  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-878-474-6 -   61.00  112.65  138.09    3655  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-735-545-14 -   60.50  134.17  8.74      507  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-449-083-14 -   60.50  134.17  8.74      507  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-868-373-3 +   59.50  119.20  59.64     1479  !
/cgn2_6/ptodata/2/ina/backfiles1.seq:5225348-2 +   59.50  117.23  76.72     1753  ! Pate
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-484-101B-49 +   59.50  113.58  122.59    2405  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-714-524D-49 +   59.50  113.58  122.59    2405  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-058-489-19 +   59.50  106.77  293.55    4335  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-058-489-20 +   59.50  105.28  355.31    4931  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-058-489-21 +   59.50  102.13  532.16    6476  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-961-522-3 +   59.00  114.98  102.41    1903  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-217-438-3 +   59.00  114.98  102.41    1903  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-321-978-3 +   59.00  114.98  102.41    1903  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-710-584-3 +   59.00  114.98  102.41    1903  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-564-805-223 +   59.00  110.08  191.98    2908  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-121-713D-59 -   59.00  107.74  259.10    3560  !
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-835-268-59 -   59.00  107.74  259.10    3560  !
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-060-692-59 -   59.00  107.74  259.10    3560  !
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-833-391-59 -   59.00  107.74  259.10    3560  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-060-610-59 -   59.00  107.74  259.10    3560  !
/cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-10151A-59 -   59.00  107.74  259.10    356
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-007-999-1 +   59.00  105.14  361.87    4460  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-210-361-1 +   59.00  105.14  361.87    4460  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-976-259-3 -   59.00   97.35  982.80    8752
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-238-481-1 +   58.50  124.58  29.91      741  !
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-247-373B-47 +   58.50  122.87  37.23      859
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-197-063-1 +   58.50  116.04  89.38     1551
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-700-749A-8 -   58.50  112.21  146.03    2160

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-332-576-1

seq_documentation_block:
; Sequence 1, Application US/08332576
; Patent No. 5756105
;  GENERAL INFORMATION:
;    APPLICANT:  Weiser, Jeffrey M.
;    TITLE OF INVENTION:  Vaccines for Haemophilis Influenza
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz and No. 5756105ris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  U.S.A.
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,576
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ralph, Rebecca L.
;      REGISTRATION NUMBER:  35,152
;      REFERENCE/DOCKET NUMBER:  CH-536
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1553..2005
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  209..1492
US-08-332-576-1


alignment_scores:
            Quality:   67.50            Length:     42
              Ratio:  2.045              Gaps:      1
 Percent Similarity:  78.571   Percent Identity: 33.333

alignment_block:
 US-10-048-197-2 x US-08-332-576-1   ..

 Align seg 1/1  to: US-08-332-576-1  from: 1  to: 2100

      11 SerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGlyAs 27
         |||   ::::::|||||||||||||:::|||||||||||||::::::::||
    1568 TCTCTTATTTTTACCGCACTTTTAATGACTGGCTGTGTGCAAAATGCGAA 1617

      27 nValAlaMetLysGluGlnAsnGlnGlnThrIleGlu.........GlnT 41
         ||||:::      :::   :::::::::   :::|||         ::::
    1618 TGTAACAACACCTCAAGCGCAAAAAATGCAAGTAGAAAAAGTGGATAAAG 1667

      41 hrIleIleLysGlyLysThrAsnLys 49
```

```
        :::::   ||||||:::::::::::::
    1668 CCTTACAAAAAGGCGAAGCTGATCGA 1693

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-13672-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9513672
;  GENERAL INFORMATION:
;    APPLICANT:  Weiser, Jeffrey M.
;    TITLE OF INVENTION:  Vaccines for Haemophilis Influenza
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz and
;      ADDRESSEE:  Norris
;      STREET:  One Liberty Place - 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  U.S.A.
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/13672
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/332,576
;      FILING DATE:  October 31, 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ralph, Rebecca L.
;      REGISTRATION NUMBER:  35,152
;      REFERENCE/DOCKET NUMBER:  CH-536
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2100 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1553..2005
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  209..1492
PCT-US95-13672-1


alignment_scores:
            Quality:   67.50            Length:     42
              Ratio:  2.045              Gaps:      1
 Percent Similarity:  78.571   Percent Identity: 33.333

alignment_block:
 US-10-048-197-2 x PCT-US95-13672-1   ..

 Align seg 1/1  to: PCT-US95-13672-1  from: 1  to: 2100

      11 SerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGlyAs 27
         |||   ::::::||||||||||||:::||||||||||||:::::::::||
    1568 TCTCTTATTTTTACCGCACTTTTAATGACTGGCTGTGTGCAAAATGCGAA 1617

      27 nValAlaMetLysGluGlnAsnGlnGlnThrIleGlu.........GlnT 41
         ||||:::      :::   :::::::::   :::|||        ::::
    1618 TGTAACAACACCTCAAGCGCAAAAAATGCAAGTAGAAAAAGTGGATAAAG 1667

      41 hrIleIleLysGlyLysThrAsnLys 49
```

```
        :::::   ||||||:::::::::::::
    1668 CCTTACAAAAAGGCGAAGCTGATCGA 1693

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-023-082A-23

seq_documentation_block:
; Sequence 23, Application US/09023082A
; Patent No. 6077692
;  GENERAL INFORMATION:
;    APPLICANT:  RUBEN, STEVEN M.
;    APPLICANT:  JIMENEZ, PABLO
;    APPLICANT:  DUAN, D. ROXANNE
;    APPLICANT:  RAMPY, MARK A.
;    APPLICANT:  MENDRICK, DONNA
;    APPLICANT:  ZHANG, JUN
;    APPLICANT:  NI, JIAN
;    APPLICANT:  MOORE, PAUL A.
;    APPLICANT:  COLEMAN, TIMOTHY A.
;    APPLICANT:  GRUBER, JOACHIM R.
;    APPLICANT:  DILLON, PATRICK J.
;    APPLICANT:  GENTZ, REINER L.
;    TITLE OF INVENTION:  KERATINOCYTE GROWTH FACTOR-2
;    NUMBER OF SEQUENCES:  148
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;      STREET:  1100 NEW YORK AVE, NW, SUITE 600
;      CITY:  WASHINGTON
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/023,082A
;      FILING DATE:  13-FEB-1998
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/01790
;      FILING DATE:  14-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/461,195
;      FILING DATE:  05-JUN-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/023,852
;      FILING DATE:  13-AUG-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/039,045
;      FILING DATE:  28-FEB-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/862,432
;      FILING DATE:  23-MAY-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/910,875
;      FILING DATE:  13-AUG-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/055,561
;      FILING DATE:  13-AUG-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  STEFFFE, ERIC K.
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.0360008/EKS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-371-2600
;      TELEFAX:  202-371-2540
;  INFORMATION FOR SEQ ID NO:  23:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4177 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  both
```

```
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  593..1216
US-09-023-082A-23


alignment_scores:
            Quality:   67.50            Length:    100
              Ratio:  1.298              Gaps:      4
 Percent Similarity:  52.000   Percent Identity: 28.000

alignment_block:
 US-10-048-197-2 x US-09-023-082A-23/rev   ..

 Align seg 1/1  to reverse of: US-09-023-082A-23  from: 1  to: 4177

      19 ValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGl 35
         |||||||||      :::|||   |||          |||:::|||:::|| 
     699 GTGACAGGGACGGAAGACACCAAGAACAGCAACAAAAAGCAGCAGCAGCA 650

      35 nGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluI 52
         ||||      :::               ||||||:::::   |||   :  
     649 GCAGCCGGGCAGGTGG............GGAAAGGCTGAGGCACAATGTG 612

      52 leSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValVal 68
         ::|||   :::|||      :::   :::::|||   :::   :::|||  
     611 TCAGTATCCATTTCCACATTGTACTGAAACTCTCGGCACTGGAAATTGTC 562

      69 IleLysPheGly.HisThrAlaIleLeuAlaPro.............. 79
            :::   ||| ||||||:::   :::   |||               
   ° 561 TCATCAGAAGGAACATACTGGAAGGGTAAGACCCGATGCAAGGCAAGGAG 512

      80 .....AsnArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrp 94
              :::||||||              |||:::::|||   ||||||  
     511 AGAGCTCGAGGTGGTGG............CTGCTGGTTAGCTCCCTCTGG 474

      95 ValLysProTyrArgProLysAsnLeuSerPheTyrLeuThrAlaLys 110
         ::::::   :::          |||          :::|||   :::  
     473 GCGCGGATCTGGCCAGAAGTGAATGCACCAACATCCATAACTCCTCGG 426

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-988-856B-3

seq_documentation_block:
; Sequence 3, Application US/08988856B
; Patent No. 6291642
; GENERAL INFORMATION:
;  APPLICANT: Weinstein, Jasminder
;  TITLE OF INVENTION: No. 6291642el Mammalian Cell Cycle Protein
;  FILE REFERENCE: 06843.0026-04; A-283 D
;  CURRENT APPLICATION NUMBER: US/08/988,856B
;  CURRENT FILING DATE:  1997-12-11
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1700
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-08-988-856B-3


alignment_scores:
            Quality:   66.50            Length:     64
              Ratio:  1.583              Gaps:      1
 Percent Similarity:  65.625   Percent Identity: 26.562

alignment_block:
 US-10-048-197-2 x US-08-988-856B-3/rev   ..

 Align seg 1/1  to reverse of: US-08-988-856B-3  from: 1  to: 1700

      17 LeuLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGl 33
         |||::::::      |||::::::   :::   ::::::::::::::::  
    1181 TTGATGCTGGGTGAATGTCTGCAGAGGAACCCAGCCACCCTCTCCAGGAG 1132

      33 nAsnGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysG 50
         :::::::      |||:::|||:::||||||         |||:::  
    1131 CACTAGGCCACACATTGACCAAGTTATCATT.........ACCACCACTG 1091

      50 lnGluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIle 66
           ::::::||||||   :::|||   |||:::         |||:::  
    1090 GCCAAATGTCGTCCATCTGGGGCCCAGCGCAGCCCACACACTTCCTGGCT 1041

      67 ValValIleLysPheGlyHisThrAlaIleLeuAlaProAsn 80
         |||:::   :::   ||||||   ::::::      ||||||  
    1040 GTGGCCACTCAGTGTGGCCACATGGTGTTCTGCTACCCGAAC 999

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-01806-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9501806
;  GENERAL INFORMATION:
;    APPLICANT:  Weinstein, Jasminder
;    TITLE OF INVENTION:  Novel Mammalian Cell Cycle Protein
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amgen Inc./Patent Operations/RBW
;      STREET:  1840 Dehavilland Drive
;      CITY:  Thousand Oaks
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  91320-1789
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/01806
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1616
PCT-US95-01806-3


alignment_scores:
            Quality:   66.50            Length:     64
              Ratio:  1.583              Gaps:      1
 Percent Similarity:  65.625   Percent Identity: 26.562

alignment_block:
 US-10-048-197-2 x PCT-US95-01806-3/rev   ..

 Align seg 1/1  to reverse of: PCT-US95-01806-3  from: 1  to: 1700

      17 LeuLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGl 33
         |||::::::      |||::::::   :::   ::::::::::::::::  
    1181 TTGATGCTGGGTGAATGTCTGCAGAGGAACCCAGCCACCCTCTCCAGGAG 1132

      33 nAsnGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysG 50
         :::::::      |||:::|||:::||||||         |||:::  
    1131 CACTAGGCCACACATTGACCAAGTTATCATT.........ACCACCACTG 1091

      50 lnGluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIle 66
           ::::::||||||   :::|||   |||:::         |||:::  
    1090 GCCAAATGTCGTCCATCTGGGGCCCAGCGCAGCCCACACACTTCCTGGCT 1041
```

```
     67 ValValIleLysPheGlyHisThrAlaIleLeuAlaProAsn 80
        |||:::   :::   ||||||   ::::::      ||||||
   1040 GTGGCCACTCAGTGTGGCCACATGGTGTTCTGCTACCCGAAC 999

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-196-520-5

seq_documentation_block:
; Sequence 5, Application US/09196520
; Patent No. 6204039
; GENERAL INFORMATION:
;  APPLICANT: Falco, Carl S.
;  APPLICANT: Allen, Stephen M.
;  TITLE OF INVENTION: Plant Isocitrate Dehydrogenase Homologs
;  FILE REFERENCE: BB-1291
;  CURRENT APPLICATION NUMBER: US/09/196,520
;  CURRENT FILING DATE: 1998-11-19
;  EARLIER APPLICATION NUMBER: 60/067,388
;  EARLIER FILING DATE: December 2, 1997
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1402
;   TYPE: DNA
;   ORGANISM: Glycine max
US-09-196-520-5


alignment_scores:
            Quality:   66.00             Length:    109
              Ratio:  1.100               Gaps:      6
 Percent Similarity:  55.046   Percent Identity: 26.606

alignment_block:
 US-10-048-197-2 x US-09-196-520-5   ..

 Align seg 1/1  to: US-09-196-520-5  from: 1  to: 1402

    . 15 ThrAlaLeuLeuValThrGlyCysValSerThrGlyAsnValAlaMet.. 30
        ||||||||||||:::::|||   :::         ::::::   :::
     993 ACTGCTTTACTGCTAAGTGGTGTTACAATGTTGCGCCATTTGAACCTCCA 1042

      31 .......LysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLysG 45
                 ||||||   |||:::::|||   :::|||||   :::|
    1043 TGACAAAGCAGAACAGATTCAAAAAGCCATCCTGAACACAATTGCAGAAG 1092

      45 lyLysThrAsnLysGlnGluIleSerSerArgPheGlySerAlaAspSer 61
         |||||   :::      :::::::::         |||||||:::   :::
    1093 GGAAGTACCGAACAGCTGACCTCGGT.........GGCAGTTCAAAGACA 1133

      62 IleSerPheMetIleValValIleLysPheGlyHisThrAlaIleLeuAl 78
           :::|||      :::::::|||         |||      ||||||
    1134 ACTGAATTCACAAAGGCAATTATT......GATCATCTTTAGATTTTG.. 1175

      78 aProAsnArgTrpGlnGluIleLeuSer..................... 87
                        :::|||:::
    1176 ..................GTTTTGGAGTCAGCAAGAAATTCTTGCTTAT 1206

      88 ........LeuIleIleSerPheLeuTrpValLysProTyrArgProLys 101
                 :::   ::::::||||||:::::::::|||   :::|||||||
    1207 GATTTTGATTTTCAGTCAATTTTCTATTTATTAACCCAATTAATCCGAAA 1256

     102 AsnLeuSerPheTyrLeuThrAlaLys 110
         :::   :::   ||||||:::   |||
    1257 GAGTGGGAT...TACCTCAGCTTAAAG 1280

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US96-05320A-333

seq_documentation_block:
; Sequence 333, Application PC/TUS9605320A
;  GENERAL INFORMATION:
;    APPLICANT:  Human Genome Sciences
;    APPLICANT:  9410 Key West Avenue
;    APPLICANT:  Rockville, MD 20850
;    APPLICANT:  United States of America
;    APPLICANT:  Johns Hopkins University
;    APPLICANT:  720 Rutland Avenue
;    APPLICANT:  Baltimore, MD 21205
;    APPLICANT:  United States of America
;    APPLICANT:  Mark D. Adams
;    APPLICANT:  Owen White
;    APPLICANT:  Hamilton O. Smith
;    APPLICANT:  J. Craig Venter
;    TITLE OF INVENTION:  Nucleotide Sequence of the Haemophilus Influenzae Rd Genome
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20003-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/05320A
;      FILING DATE:  April22, 1996
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/476,102
;      FILING DATE:  June 7, 1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/487,429
;      FILING DATE:  June 7, 1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Eric K. Steffe
;      REGISTRATION NUMBER:  36,688
;      REFERENCE/DOCKET NUMBER:  1488.014PC01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  333:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  402 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
PCT-US96-05320A-333


alignment_scores:
            Quality:   65.50             Length:     42
              Ratio:  2.047               Gaps:      1
 Percent Similarity:  76.190   Percent Identity: 33.333

alignment_block:
 US-10-048-197-2 x PCT-US96-05320A-333   ..

 Align seg 1/1  to: PCT-US96-05320A-333  from: 1  to: 402

     11 SerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGlyAs 27
        |||   ::::::||||||||||||:::|||||||||||:::::::::||
     16 TCTCTTATTTTTACCGCACTTTTAATGACTGGCTGTGTGCAAAATGCGAA 65

     27 nValAlaMetLysGluGlnAsnGlnGlnThrIleGlu.........GlnT 41
        ||||:::      :::   :::::::::   :::|||         ::::
     66 TGTAACAACACCTCAAGCGCAAAAAATGCAAGTAGAAAAAGTGGATAAAG 115

     41 hrIleIleLysGlyLysThrAsnLys 49
        :::::   ||||||::::::   :::
    116 CCTTACAAAAAGGCGAAGCTGNTCGA 141

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-371-377-16
```

```
seq_documentation_block:
; Sequence 16, Application US/08371377
; Patent No. 5851764
;  GENERAL INFORMATION:
;    APPLICANT:  Fisher, Paul B.
;    APPLICANT:  Shen, Ruoqian
;    TITLE OF INVENTION:  DEVELOPMENT OF DNA PROBES AND
;    TITLE OF INVENTION:  IMMUNOLOGICAL REAGENTS SPECIFIC FOR CELL SURFACE-EXPRESSED
;    TITLE OF INVENTION:  MOLECULES AND TRANSFORMATION-ASSOCIATED GENES
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/371,377
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0575/37590-B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2128 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  621..1817
US-08-371-377-16


alignment_scores:
            Quality:   64.50            Length:    119
              Ratio:  1.057              Gaps:      5
 Percent Similarity:  51.261   Percent Identity: 22.689

alignment_block:
 US-10-048-197-2 x US-08-371-377-16   ..

 Align seg 1/1  to: US-08-371-377-16  from: 1  to: 2128

      18 LeuValThrGlyCys..............ValSerThrGlyAsnValAl 29
         |||::::::   |||              :::||||||      :::::
     646 TTGATATCTCCTTGTGGAAATTTGAGACCAGCAAGTACTATGTGACTATC 695

      29 aMetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLysGlyL 46
         ::::      |||:::   :::|||:::::::::|||   :::::::|||
     696 ATTGATGCCCCAGGACACAGAGACTTTATCCAAAACATGATTACAGGGAC 745

      46 ys.....................................ThrAsn 48
                                                :::|||
     746 CTCTCAGGCTGACTGTGCTGTCCTGATTGTTGCTGCTGGTGTTGGTGAAT 795

      49 LysGlnGluIleSerSerArgPheGlySerAlaAspSerIleSerPheMe 65
            :::   :::|||   |||:::|||      ||||||:::   |||
     796 TTGAAGCTGGTATCTCCAAGAATGGGCAGACCCGACAGCATGCCCTTCTG 845

      65 tIle.............................ValValIleLysPhe. 71
            :::                          |||:::   |||:::
     846 GCTTACACACTGGGTGTGAAACAACTAATTGTCGGTGTTAACAAAATGGA 895

      72 GlyHisThrAlaIleLeuAlaPro.AsnArgTrpGlnGluIleLeuSerL 88
            |||   |||   |||   ||| :::|||::::::||||||::::::
     896 TTCCACTGAGCCACCCTACAGCCAGAAGAGATATGAGGAAATTGTTAAGG 945

      88 euIleIleSerPheLeuTrpValLysProTyrArgProLysAsnLeuSer 104
            :::   :::::::::           |||:::|||   :::::::::
     946 AAGTCAGCACTTACATTAAGAAAATTGGCTACAACCCCGACACAGTAGCA 995

     105 Phe 105
         |||
     996 TTT 998

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-895-300A-2

seq_documentation_block:
; Sequence 2, Application US/07895300A
; Patent No. 5279823
;  GENERAL INFORMATION:
;    APPLICANT:  Frenz, John
;    APPLICANT:  Shire, Steven J.
;    APPLICANT:  Sliwkowski, Mary B.
;    TITLE OF INVENTION:  PURIFIED FORMS OF DNase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/895,300A
;      FILING DATE:  19920608
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hensley, Max D.
;      REGISTRATION NUMBER:  27,043
;      REFERENCE/DOCKET NUMBER:  747
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1994
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1039 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-895-300A-2


alignment_scores:
            Quality:   63.50            Length:    124
              Ratio:  0.977              Gaps:      6
 Percent Similarity:  52.419   Percent Identity: 24.194

alignment_block:
 US-10-048-197-2 x US-07-895-300A-2   ..

 Align seg 1/1  to: US-07-895-300A-2  from: 1  to: 1039
```

```
      3 AsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuVa 19
        :::||||||:::|||:::             ||||||   |||||||||:: 
    145 CACCATCATCTCAGGATGAGGGGCATGAAGCTGCTGGGGGCGCTGCTGGC 194

     19 lThrGlyCysValSerThrGlyAsnValAlaMetLys......GluGlnA 34
        :   :::::::::     |||   |||::::::|||            | 
    195 ACTGGCGGCCCTACTGCAGGGGGCCGTGTCCCTGAAGATCGCAGCCTTCA 244

     34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
        ||   ||||||:::   :::|||   :::::::::   |||      ::: 
    245 ACATCCAGACATTTGGGGAGACCAAGATGTCCAATGCCACCCTCGTCAGC 294

     51 .........GluIleSerSerArgPheGlySerAlaAspSerIleSerPh 64
                :::|||   ||||||:::      |||                
    295 TACATTGTGCAGATCCTGAGCCGCTATGACATCGCC.............. 330

     64 eMetIleValValIleLysPheGlyHisThrAlaIleLeu.......... 77
         ::::::      ::::::   :::|||   :::   :::           
    331 .CTGGTCCAGGAGGTCAGAGACAGCCACCTGACTGCCGTGGGGAAGCTGC 379

     78 ....................AlaProAsnArgTrpGlnGluIleLeuSer 87
                            ||||||:::   ::::::   ::::::||| 
    380 TGGACAACCTCAATCAGGATGCACCAGACACCTATCACTACGTGGTCAGT 429

     88 LeuIleIleSer.....................PheLeuTrpValLysPr 97
           ::::::                       :::|||:::||| 
    430 GAGCCACTGGGACGGAACAGCTATAAGGAGCGCTACCTGTTCGTG..... 474

     97 oTyrArgProLysAsnLeuSer 104
         |||||||||   ::::::||| 
    475 .TACAGGCCTGACCAGGTGTCT 495

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-367-2

seq_documentation_block:
; Sequence 2, Application US/08458367
; Patent No. 5783433
;  GENERAL INFORMATION:
;    APPLICANT:  Frenz, John
;    APPLICANT:  Shire, Steven J.
;    APPLICANT:  Sliwkowski, Mary B.
;    TITLE OF INVENTION:  PURIFIED FORMS OF DNase
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,367
;      FILING DATE:  02-Jun-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/409631
;      FILING DATE:  22-Mar-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/348284
;      FILING DATE:  30-No. 5783433-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/116186
;      FILING DATE:  02-Sep-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/895300
;      FILING DATE:  08-Jun-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Johnston, Sean A.
;      REGISTRATION NUMBER:  35,910
;      REFERENCE/DOCKET NUMBER:  P0747C4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3562
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1039 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-458-367-2


alignment_scores:
           Quality:   63.50             Length:    124
             Ratio:  0.977               Gaps:      6
 Percent Similarity:  52.419   Percent Identity: 24.194

alignment_block:
 US-10-048-197-2 x US-08-458-367-2   ..

 Align seg 1/1  to: US-08-458-367-2  from: 1  to: 1039

      3 AsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuVa 19
        :::||||||:::|||:::             ||||||   |||||||||:: 
    145 CACCATCATCTCAGGATGAGGGGCATGAAGCTGCTGGGGGCGCTGCTGGC 194

     19 lThrGlyCysValSerThrGlyAsnValAlaMetLys......GluGlnA 34
        :   :::::::::     |||   |||::::::|||            | 
    195 ACTGGCGGCCCTACTGCAGGGGGCCGTGTCCCTGAAGATCGCAGCCTTCA 244

     34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
        ||   ||||||:::   :::|||   :::::::::   |||      ::: 
    245 ACATCCAGACATTTGGGGAGACCAAGATGTCCAATGCCACCCTCGTCAGC 294

     51 .........GluIleSerSerArgPheGlySerAlaAspSerIleSerPh 64
                :::|||   ||||||:::      |||                
    295 TACATTGTGCAGATCCTGAGCCGCTATGACATCGCC.............. 330

     64 eMetIleValValIleLysPheGlyHisThrAlaIleLeu.......... 77
         ::::::      ::::::   :::|||   :::   :::           
    331 .CTGGTCCAGGAGGTCAGAGACAGCCACCTGACTGCCGTGGGGAAGCTGC 379

     78 ....................AlaProAsnArgTrpGlnGluIleLeuSer 87
                            ||||||:::   ::::::   ::::::||| 
    380 TGGACAACCTCAATCAGGATGCACCAGACACCTATCACTACGTGGTCAGT 429

     88 LeuIleIleSer.....................PheLeuTrpValLysPr 97
           ::::::                       :::|||:::||| 
    430 GAGCCACTGGGACGGAACAGCTATAAGGAGCGCTACCTGTTCGTG..... 474

     97 oTyrArgProLysAsnLeuSer 104
         |||||||||   ::::::||| 
    475 .TACAGGCCTGACCAGGTGTCT 495

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-05136-2

seq_documentation_block:
; Sequence 2, Application PC/TUS9305136
;  GENERAL INFORMATION:
;    APPLICANT:  Genentech, Inc.
;    TITLE OF INVENTION:  PURIFIED FORMS OF DNase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/05136
;      FILING DATE:  19930528
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Johnston, Sean A.
;      REGISTRATION NUMBER:  P35,910
;      REFERENCE/DOCKET NUMBER:  747PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3562
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1039 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US93-05136-2


alignment_scores:
            Quality:   63.50              Length:    124
              Ratio:  0.977                Gaps:      6
 Percent Similarity:  52.419   Percent Identity: 24.194

alignment_block:
 US-10-048-197-2 x PCT-US93-05136-2   ..

 Align seg 1/1  to: PCT-US93-05136-2  from: 1  to: 1039

       3 AsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuVa 19
         :::|||||||:::|||:::            ||||||   |||||||||::
     145 CACCATCATCTCAGGATGAGGGGCATGAAGCTGCTGGGGGCGCTGCTGGC 194

      19 lThrGlyCysValSerThrGlyAsnValAlaMetLys......GluGlnA 34
         :   :::::::::      |||   |||:::::::|||            |
     195 ACTGGCGGCCCTACTGCAGGGGGCCGTGTCCCTGAAGATCGCAGCCTTCA 244

      34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
         ||   ||||||:::   :::|||   :::::::::   |||      :::
     245 ACATCCAGACATTTGGGGAGACCAAGATGTCCAATGCCACCCTCGTCAGC 294

      51 .........GluIleSerSerArgPheGlySerAlaAspSerIleSerPh 64
                 :::|||   ||||||:::      |||
     295 TACATTGTGCAGATCCTGAGCCGCTATGACATCGCC.............. 330

      64 eMetIleValValIleLysPheGlyHisThrAlaIleLeu.......... 77
          ::::::      ::::::   :::|||   :::   :::
     331 .CTGGTCCAGGAGGTCAGAGACAGCCACCTGACTGCCGTGGGGAAGCTGC 379

      78 ....................AlaProAsnArgTrpGlnGluIleLeuSer 87
                             ||||||:::   ::::::   ::::::|||
     380 TGGACAACCTCAATCAGGATGCACCAGACACCTATCACTACGTGGTCAGT 429

      88 LeuIleIleSer.....................PheLeuTrpValLysPr 97
               ::::::                    :::|||:::|||
     430 GAGCCACTGGGACGGAACAGCTATAAGGAGCGCTACCTGTTCGTG..... 474

      97 oTyrArgProLysAsnLeuSer 104
          |||||||||   :::::::|||
     475 .TACAGGCCTGACCAGGTGTCT 495

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-10519-1


seq_documentation_block:
; Sequence 1, Application PC/TUS9310519
;  GENERAL INFORMATION:
;    APPLICANT:  Genentech, Inc.
;    TITLE OF INVENTION:  Compaction Assay for Assessment of Respiratory Disease
;    TITLE OF INVENTION:  Therapy
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/10519
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Johnston, Sean A.
;      REGISTRATION NUMBER:  35,910
;      REFERENCE/DOCKET NUMBER:  792C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3562
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1039 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US93-10519-1


alignment_scores:
            Quality:   63.50              Length:    124
              Ratio:  0.977                Gaps:      6
 Percent Similarity:  52.419   Percent Identity: 24.194

alignment_block:
 US-10-048-197-2 x PCT-US93-10519-1   ..

 Align seg 1/1  to: PCT-US93-10519-1  from: 1  to: 1039

       3 AsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuVa 19
         :::|||||||:::|||:::            ||||||   |||||||||::
     145 CACCATCATCTCAGGATGAGGGGCATGAAGCTGCTGGGGGCGCTGCTGGC 194

      19 lThrGlyCysValSerThrGlyAsnValAlaMetLys......GluGlnA 34
         :   :::::::::      |||   |||:::::::|||            |
     195 ACTGGCGGCCCTACTGCAGGGGGCCGTGTCCCTGAAGATCGCAGCCTTCA 244

      34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
         ||   ||||||:::   :::|||   :::::::::   |||      :::
     245 ACATCCAGACATTTGGGGAGACCAAGATGTCCAATGCCACCCTCGTCAGC 294

      51 .........GluIleSerSerArgPheGlySerAlaAspSerIleSerPh 64
                 :::|||   ||||||:::      |||
     295 TACATTGTGCAGATCCTGAGCCGCTATGACATCGCC.............. 330

      64 eMetIleValValIleLysPheGlyHisThrAlaIleLeu.......... 77
          ::::::      ::::::   :::|||   :::   :::
     331 .CTGGTCCAGGAGGTCAGAGACAGCCACCTGACTGCCGTGGGGAAGCTGC 379
```

```
      78 ..................AlaProAsnArgTrpGlnGluIleLeuSer 87
                           ||||||:::   ::::::   :::::::||
     380 TGGACAACCTCAATCAGGATGCACCAGACACCTATCACTACGTGGTCAGT 429

      88 LeuIleIleSer.....................PheLeuTrpValLysPr 97
            ::::::                       :::|||:::|||
     430 GAGCCACTGGGACGGAACAGCTATAAGGAGCGCTACCTGTTCGTG..... 474

      97 oTyrArgProLysAsnLeuSer 104
          |||||||||   ::::::|||
     475 .TACAGGCCTGACCAGGTGTCT 495
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-013-810-1

```
seq_documentation_block:
; Sequence 1, Application US/09013810
; Patent No. 6197551
;  GENERAL INFORMATION:
;    APPLICANT:  Busfield, Samantha J.
;    TITLE OF INVENTION:  No. 6197551el TANGO 80 Molecules and Uses Thereof
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/013,810
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Amy E. Mandragouras
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  MEI-010
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  746 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  135..428
US-09-013-810-1


alignment_scores:
            Quality:   63.00             Length:    111
              Ratio:  1.050               Gaps:      5
 Percent Similarity:  54.054   Percent Identity: 24.324

alignment_block:
 US-10-048-197-2 x US-09-013-810-1   ..

 Align seg 1/1  to: US-09-013-810-1  from: 1  to: 746

       4 HisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValTh 20
         ||||||:::||||||   :::      :::|||      :::::::|||
     117 CATCACCTTCGCTTAGACATGTTCAGGATCTTA......GTAGTCGTGTG 160
```

```
      20 rGly...CysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnG 36
         |||   |||   :::   :::::::::      :::  :::::::::|
     161 TGGATCCTGCAGAACAATATCCTCACTGCAGTCCCAAGGAAAGAGCAAAC 210

      36 lnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluIle 52
         ||      |||   :::|||::::::      ::::::|||:::   :::
     211 AGTTCCAGGAAGGGAACATAATGGAAATGTACAACAAAAAGGAACCTGTA 260

      53 ................................SerSerArgPheGl 57
                                         :::|||   |||
     261 AAAGCCTCTCTCTTCTATCACAAGAAGAGTGGTACAACCTCTACATTTGA 310

      57 ySerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHisT 74
         ||||||        :::::::::|||:::|||   :::   |||
     311 GTCTGCAGCCTTCCCTGGTTGGTTCATCGCTGTCTGCTCTAAAGGGAGCT 360

      74 hrAlaIleLeuAla.ProAsnArgTrpGlnGluIleLeuSerLeuIleIl 90
            ::::::    |||:::::|||   :::      ||||||
     361 GCCCACTCATTCTGACCCAAGAACTGGGGGAAATCTTCATCACTGACTTC 410

      90 eSerPheLeuTrp...ValLysProTyrArg 99
            ||||||   :::|||   :::|||
     411 GAGATGATTGTGGTACATTAAGGTTTTTAGA 441
```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-058-489-34

```
seq_documentation_block:
; Sequence 34, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
;  APPLICANT: Whitehead Institute for Biomedical Research
;  APPLICANT: Lahn, Bruce
;  APPLICANT: Page, David
;  TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
;  TITLE OF INVENTION: the Y Chromosome
;  FILE REFERENCE: WHI97-08pA
;  CURRENT APPLICATION NUMBER: US/09/058,489
;  CURRENT FILING DATE: 1998-04-10
;  EARLIER APPLICATION NUMBER: 60/041,877
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 91
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 10091
;   TYPE: DNA
;   ORGANISM: Human
US-09-058-489-34


alignment_scores:
            Quality:   63.00             Length:    103
              Ratio:  1.086               Gaps:      4
 Percent Similarity:  56.311   Percent Identity: 23.301

alignment_block:
 US-10-048-197-2 x US-09-058-489-34   ..

 Align seg 1/1  to: US-09-058-489-34  from: 1  to: 10091

      10 IleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGl 26
         |||   |||:::         :::   ::::::::::|||   :::   ||
    5382 ATTAGGGCTATACAGAAAATTATCTGGGCATCAGCATGTGGGGCATTAGG 5431

      26 yAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleI 43
         |   |||         :::::::::          |||   |||
    5432 ACTAGTTTTTAGCCCAAATGAAGAAATAACTAAAATTTATCAGATGACCA 5481

      43 leLysGlyLysThrAsnLysGlnGluIleSerSerArgPheGlySerAla 59
         :::|||   :::||||||   |||:::::::::::::         :::
    5482 CCAATGGA...AGCAATAAGCTGGAGGTGGAAGATGAACAAGTTTGCTGT 5528

      60 AspSerIleSerPheMetIleValValIleLysPheGlyHisThrAlaIl 76
```

```
          :::::::::::::   |||   :::   :::
      5529 GAAGCACTGGAAGTGATGACCTTATGTTTT.................GC 5560

        76 eLeuAlaPro.AsnArgTrpGlnGluIleLeuSerLeuIleIleSerPhe 92
           |||   ||| :::|||||||:::::::::::::::
      5561 TTTACTTCCAACAGCGTTGGATGCACTTAGTAAAG.............. 5595

        93 LeuTrpValLysProTyrArgProLysAsnLeuSerPheTyrLeuThrAl 109
                  ||||||   ||||||:::::::|||:::::::|||   |||::
      5596 ......AAAAAGCCTGGCAGACCTTCATCATTGACTTATTATTGCACTGT 5639

       109 aLysAla 111
           ::::|||
      5640 CCAAGCA 5646

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-366-783-3

seq_documentation_block:
; Sequence 3, Application US/08366783
; Patent No. 5650554
;  GENERAL INFORMATION:
;    APPLICANT:  Moloney, Maurice M
;    TITLE OF INVENTION:  Oil-Body Proteins As Carriers Of
;    TITLE OF INVENTION:  High-Value Peptides In Plants
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DEHLINGER & ASSOCIATES
;      STREET:  350 CAMBRIDGE AVENUE, SUITE 250
;      CITY:  PALO ALTO
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  94025-1536
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/366,783
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FABIAN, GARY
;      REGISTRATION NUMBER:  33,875
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-324-0880
;      TELEFAX:  415-324-0960
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  765 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-366-783-3


alignment_scores:
            Quality:   61.50             Length:     98
              Ratio:  1.098               Gaps:      4
 Percent Similarity:  57.143   Percent Identity: 23.469

alignment_block:
 US-10-048-197-2 x US-08-366-783-3   ..

 Align seg 1/1  to: US-08-366-783-3  from: 1  to: 765

        8 LeuThrIleSerAlaLeuLeuThr...AlaLeuLeuValThrGlyCysVa 23
          :::   :::   ||||||:::|||   |||||||||:::||||||   ::
      460 ATCCTCGTGCCGGCTCTCATCACCGTAGCACTTCTCATCACTGGCTTTCT 509

       23 lSerThrGlyAsnValAlaMet.......................... 30
          :|||:::|||:::   |||:::

      510 CTCCTCTGGTGGGTTTGCCATTGCAGCTATAACCGTCTTCTCCTGGATCT 559

       31 ...................LysGluGlnAsnGlnGlnThrIleGluGln 40
                                 |||::::::::::   :::::::::
      560 ATAAGTACGCAACGGGAGAGCACCCACAGGGGTCAGATAAGTTGGACAGT 609

       41 ThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArg.PheG 57
          :::   :::|||      |||:::   |||:::|||::::::||| ::::
      610 GCAAGGATGAAGCTGGGAACCAAAGCTCAGGATATTAAAGACAGAGCTCA 659

       57 lySerAlaAspSer............IleSerPheMetIleValValIle 69
          ::::::::||||||           :::|||:::   ::::::|||:::
      660 ATACTACGGACAGCAACATACAGGTGGTGAGCATGACCGTGACCGTACTC 709

       70 LysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 83
             :::   |||   :::         :::   |||:::
      710 GTGGTGGCCAGCACACTACTATCGAAGGAAGAGCCATGGCGC 751

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-514-302-1

seq_documentation_block:
; Sequence 1, Application US/09514302
; Patent No. 6338959
; GENERAL INFORMATION:
;  APPLICANT: HATADA, Yuji
;  APPLICANT: IGARASHI, Kazuaki
;  APPLICANT: OZAKI, Katsuya
;  APPLICANT: ARA, Katsutoshi
;  APPLICANT: KAWAI, Shuji
;  APPLICANT: ITO, Susumu
;  TITLE OF INVENTION: GENE FOR ENZYME HAVING BOTH ALKALINE PULLULANASE AND
;  TITLE OF INVENTION: ALKALINE ALPHA-AMYLASE ACTIVITIES
;  FILE REFERENCE: 2173-105P
;  CURRENT APPLICATION NUMBER: US/09/514,302
;  CURRENT FILING DATE: 2000-02-28
;  EARLIER APPLICATION NUMBER: 08/952,084
;  EARLIER FILING DATE: 1997-11-10
;  NUMBER OF SEQ ID NOS: 14
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 6142
;   TYPE: DNA
;   ORGANISM: Bacillus sp.
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (145)..(5958)
;   FEATURE:
;   NAME/KEY: sig_peptide
;   LOCATION: (145)..(240)
;   FEATURE:
;   NAME/KEY: mat_peptide
;   LOCATION: (241)..(5958)
US-09-514-302-1


alignment_scores:
            Quality:   61.50             Length:    113
              Ratio:  1.118               Gaps:      4
 Percent Similarity:  48.673   Percent Identity: 21.239

alignment_block:
 US-10-048-197-2 x US-09-514-302-1   ..

 Align seg 1/1  to: US-09-514-302-1  from: 1  to: 6142

        4 HisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValTh 20
          ||||||      :::   :::   |||:::   |||   ||||||
     5672 CACCACCAGGGGATGGGGATG...GCGATGGAAACACGCCTCCTC..... 5713

       20 rGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnT 37
                  :::|||      :::   ::::::|||::::::|||:::
     5714 .......CAGGCAACGGTAATGGCAATAATCCAGGAACACCACCAGGAAA 5756
```

```
      37 hrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSer 53
         :::|||:::|||::::::|||:::||||||:::|||:::::::::::::
    5757 GGGTGGAGAAAACCCTGGTAAAGGCAAAAACGACAAAACACCGCCTGGCA 5806

      54 SerArgPheGlySerAlaAspSerIleSerPheMetIleValValIleLy 70
         :::                              :::   :::|||::
    5807 AAG..............................GTGGGGACAATCCA 5823

      70 sPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuS 87
         :      |||:::      |||:::      :::   |||   |
    5824 GGTAAGGGGAACAAGCTACCACTTACCGCAACCGGAACACTTAATTACAT 5873

      87 erLeuIleIleSerPheLeuTrpVal..................... 95
         ||   :::::::::      |||
    5874 CCTGTTTGGTGCAATAATGTTGGTTCTTGGGACGCTGCTGTATCTAGGGG 5923

      96 .........................LysProTyrArg 99
                                 |||||||||:::
    5924 TCAGAAGAAAAGCAGGATTGAAAGAAAAAACCTTATAAA 5962

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-309-487-25

seq_documentation_block:
; Sequence 25, Application US/09309487
; Patent No. 6335318
; GENERAL INFORMATION:
;  APPLICANT: Selsted, Michael E.
;  APPLICANT: Tang, Yi-Quan
;  APPLICANT: Yuan, Jun
;  APPLICANT: Ouellette, Andre J.
;  TITLE OF INVENTION: Antimicrobial Theta Defensins and Methods of Using Same
;  FILE REFERENCE: P-UC 3095
;  CURRENT APPLICATION NUMBER: US/09/309,487
;  CURRENT FILING DATE: 1999-05-10
;  NUMBER OF SEQ ID NOS: 31
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
;   LENGTH: 2553
;   TYPE: DNA
;   ORGANISM: Macaca mulatta
US-09-309-487-25


alignment_scores:
            Quality:   61.00              Length:    124
              Ratio:  0.910                Gaps:      5
 Percent Similarity:  54.032   Percent Identity: 23.387

alignment_block:
 US-10-048-197-2 x US-09-309-487-25/rev   ..

 Align seg 1/1  to reverse of: US-09-309-487-25  from: 1  to: 2553

       2 LeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLe 18
         :::|||||||:::   |||   |||   :::|||::::::|||:::
    2412 GTCAACCACGAAAAGCGCAGGACCCCAAACGCCTTTATAACAGTTGGCAA 2363

      18 uValThrGlyCysValSerThrGlyAsnValAlaMet.....L 31
          :::      |||   :::|||:::::::::::::::      :
    2362 ACTCCTCGTCTGCAAAGGCACCTCAAGCCTCTCGCTGACTCTGGGGACAC 2313

      31 ysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLysGlyLysThr 47
         ::::::::::::::   |||:::::::::   :::::   |||
    2312 AGAGAAAGCATCATAAATGGATCAGAAGGTCTGCAGTGAGTGACAAAGGG 2263

      48 AsnLysGlnGluIleSerSer.........................Ar 55
         |||      |||:::                              ||
    2262 AATCCTGGAGAAGTCATGTGCTAGCTAATGAGGGATGTCAGCTGCATTCG 2213

      55 gPheGlySerAlaAspSerIleSer..............PheMetIleV 67
         |   ||||||   ::::::::::|||                  :::|
    2212 GGCCGGTAGCATGAACAACCTCAGTCAGTAGGAATAAATACACAGAGTAG 2163
```

```
      67 alValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 83
         ||           |||||||||:::::
    2162 TGCT.........GGTCACACAGGATTT..................... 2144

      84 GluIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPr 100
         |||      |||   :::::|||||||||   :::   |||   :::
    2143 GAGACTTATTCTCATTTGCTCTCATTTTTGTGCCTGTGCCCCATCAAACA 2094

     100 oLysAsnLeuSerPheTyrLeu 107
          |||:::|||:::      |||
    2093 CAAACACCTGAATGCACTCTTA 2072

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-878-474-6

seq_documentation_block:
; Sequence 6, Application US/08878474
; Patent No. 6133232
;  GENERAL INFORMATION:
;    APPLICANT:  De Robertis, Edward M.
;    APPLICANT:  Bouwmeester, Tewis
;    TITLE OF INVENTION:  Endoderm, Cardiac and Neural Inducing
;    TITLE OF INVENTION:  Factors
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Majestic, Parsons, Siebert & Hsue
;      STREET:  Four Embarcadero Center, Suite 1100
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111-4106
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,474
;      FILING DATE:  18-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/020,150
;      FILING DATE:  20-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Siebert, J. Suzanne
;      REGISTRATION NUMBER:  28,758
;      REFERENCE/DOCKET NUMBER:  3100.002US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/248-5500
;      TELEFAX:  415/362-5418
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-878-474-6


alignment_scores:
            Quality:   61.00              Length:    100
              Ratio:  1.034                Gaps:      3
 Percent Similarity:  59.000   Percent Identity: 23.000

alignment_block:
 US-10-048-197-2 x US-08-878-474-6/rev   ..

 Align seg 1/1  to reverse of: US-08-878-474-6  from: 1  to: 3655

       3 AsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuVa 19
         |||:::|||:::   :::  ||||||   |||         :::|||::
    1367 AACTATCATGTAACTGTCCTCATAAGCTTGCTGTAGTTTAAAGTGCTCAT 1318
```

```
      19 lThr.....................GlyCysValSerThrGlyAsnValA 29
         ::::                        :::       |||   |||:::|||:
    1317 GTCCATAAAGAGTACAGCGAACTTGTCCATTAGATCCAGAGGCTCTGTCA 1268

      29 laMetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLysGly 45
         ::      :::|||:::   :::::::::             :::   :::
    1267 GTAGTGCTGATCAGAGCTATAAAGTTCTCCTTTGTGGCTGTTTCTGGAAT 1218

      46 LysThrAsnLysGlnGluIleSerSerArgPheGlySerAlaAspSerIl 62
            |||:::         |||||||:::   |||::::::|||:::
    1217 ATAGGCAACTCCTGCATTTACAGTAGTCAGAGGGGTAATAGTGATGGCTG 1168

      62 eSerPheMetIleValValIleLysPheGlyHisThrAlaIleLeuAlaP 79
          :::   :::|||:::   ||||||:::   ::::::   :::   :::
    1167 GGGT...ATTATCATTTACATCAAGTATATGAACAGTTACTTTACAAGTA 1121

      79 roAsnArgTrp......GlnGluIleLeuSerLeuIleIleSerPheLeu 93
         ::::::|||         :::||||||:::|||   ::::::|||:::
    1120 GCAGTCAGTGGGTTGGGGCCCAAATCTTGGGCTTGTACCTCAAATTCGTA 1071

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-735-545-14

seq_documentation_block:
; Sequence 14, Application US/08735545
; Patent No. 6025131
;  GENERAL INFORMATION:
;    APPLICANT:  Van Dyk, Tina K.
;    APPLICANT:  LaRossa, Robert Alan
;    TITLE OF INVENTION:  A Facile Method for
;    TITLE OF INVENTION:  Identifying Regulated
;    TITLE OF INVENTION:  Promoters
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS
;      ADDRESSEE:  AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/735,545
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA A.
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CR-9989
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  507 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      STRAIN:  dpd2090 lower
US-08-735-545-14


alignment_scores:
            Quality:   60.50             Length:     35
              Ratio:  2.327               Gaps:      1
 Percent Similarity:  74.286   Percent Identity: 42.857

alignment_block:
 US-10-048-197-2 x US-08-735-545-14/rev   ..

 Align seg 1/1  to reverse of: US-08-735-545-14  from: 1  to: 507

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         ||||||   ||||||       |||   |||:::::::::|||:::::||
     192 AGACTGAAGATTTCGAAAACTCTGCTGGCTGTAATGTTGACCTCTGCCGT 143

       23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
          |:::||||||:::   |||       |||:::|||   ||||||   ::::
     142 CGCGACCGGCTCT...GCCTACGCGGAAAACAACGCGCAGACTACCAATG 96

       40 lnThr 41
          :::::
       95 AAAGC 91

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-449-083-14

seq_documentation_block:
; Sequence 14, Application US/09449083
; Patent No. 6194159
;  GENERAL INFORMATION:
;    APPLICANT:  Van Dyk, Tina K.
;    APPLICANT:  LaRossa, Robert Alan
;    TITLE OF INVENTION:  A Facile Method for
;    TITLE OF INVENTION:  Identifying Regulated
;    TITLE OF INVENTION:  Promoters
;    NUMBER OF SEQUENCES:  28
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS
;      ADDRESSEE:  AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  DISKETTE, 3.50 INCH
;      COMPUTER:  IBM PC COMPATIBLE
;      OPERATING SYSTEM:  MICROSOFT WINDOWS 3.1
;      SOFTWARE:  MICROSOFT WORD 2.0C
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/449,083
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  FLOYD, LINDA A.
;      REGISTRATION NUMBER:  33,692
;      REFERENCE/DOCKET NUMBER:  CR-9989
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-8112
;      TELEFAX:  302-773-0164
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  507 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      STRAIN:  dpd2090 lower
US-09-449-083-14


alignment_scores:
            Quality:   60.50             Length:     35
              Ratio:  2.327               Gaps:      1
 Percent Similarity:  74.286   Percent Identity: 42.857

alignment_block:
 US-10-048-197-2 x US-09-449-083-14/rev   ..
```

```
 Align seg 1/1  to reverse of: US-09-449-083-14  from: 1  to: 507

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         ||||||   ||||||      |||   |||:::::::::|||:::::::|| 
     192 AGACTGAAGATTTCGAAAACTCTGCTGGCTGTAATGTTGACCTCTGCCGT 143

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
         |:::|||||:::   |||      |||:::|||   ||||||   ::::  
     142 CGCGACCGGCTCT...GCCTACGCGGAAAACAACGCGCAGACTACCAATG 96

      40 lnThr 41
         :::::
      95 AAAGC 91

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-868-373-3

seq_documentation_block:
; Sequence 3, Application US/08868373
; Patent No. 6307128
; GENERAL INFORMATION:
;  APPLICANT: Jaworski, Jan G.
;  APPLICANT: Post-Beittenmiller, Martha A.
;  APPLICANT: Todd, James
;  TITLE OF INVENTION: FATTY ACID ELONGASES
;  FILE REFERENCE: 07148/064001
;  CURRENT APPLICATION NUMBER: US/08/868,373
;  CURRENT FILING DATE: 1997-06-03
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 1479
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-08-868-373-3


alignment_scores:
            Quality:   59.50             Length:     87
              Ratio:  1.214               Gaps:      4
 Percent Similarity:  56.322   Percent Identity: 22.989

alignment_block:
 US-10-048-197-2 x US-08-868-373-3   ..

 Align seg 1/1  to: US-08-868-373-3  from: 1  to: 1479

      22 CysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIl 38
         |||:::   :::|||:::   |||:::   ::::::   ::::::::::: 
    1156 TGTATCCATGCGGGTGGTAGAGCGCTAATTGATGAGATGGAGAAGAATCT 1205

      38 eGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerA 55
         ::::   |||   :::      :::::::::::::::   :::      | 
    1206 TCATCTAACTCCACTAGACGTTGAGGCTTCAAGAATGACATTA...CACA 1252

      55 rgPheGlySerAlaAspSerIleSerPhe..................... 64
         ||||||||:::::::::|||   |||::: 
    1253 GGTTTGGTAATACCTCTTCGAGCTCCATTTGGTACGAGTTGGCTTACACA 1302

      65 ..............MetIle.ValValIleLysPheGlyHisThrAlaI 76
                       |||          |||   ||||||::: 
    1303 GAAGCCAAAGGAAGGATGACGAAAGGAGATAGGATTTGGCAGATTG.... 1348

      76 leLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIleSerPhe 92
                     ||||||   ::::::|||||||:::|||::::::||| 
    1349 ..............CGTTGGGGTCAGGTTTTAAGTGTAATAGTTCAGTTT 1384

      93 LeuTrpVal 95
            |||::: 
    1385 GGGTGGCTC 1393

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5225348-2

seq_documentation_block:

;Patent No. 5225348
;    APPLICANT: HAGATA, SHIGEKAZU;SUGANO, SUMIO;KIM, DONG W.;
;UETSUKI, TAICHI;KAZIRO, YOSHITO
;    TITLE OF INVENTION: DNA FRAGMENT AND EXPRESSION PLASMID
;CONTAINING THE DNA FRAGMENT
;    NUMBER OF SEQUENCES: 9
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/447,823
;      FILING DATE: 08-DEC-1989
;SEQ ID NO:2:
;      LENGTH: 1753
5225348-2


alignment_scores:
            Quality:   59.50             Length:    127
              Ratio:  0.902               Gaps:      5
 Percent Similarity:  51.969   Percent Identity: 21.260

alignment_block:
 US-10-048-197-2 x 5225348-2   ..

 Align seg 1/1  to: 5225348-2  from: 1  to: 1753

      10 IleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCys........... 22
         :::|||:::   :::::::::   |||::::::   ||| 
     256 CTGAGCGTGAACGTGGTATCACCATTGATATCTCCTTGTGGAAATTTGAG 305

      23 ....ValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrI 38
              ||||||      :::::::::      |||:::   :::|||: 
     306 ACCAGGAAGTACTATGTGACTATCATTGATGCCCCAGGACACAGAGACTT 355

      38 leGluGlnThrIleIleLysGlyLys...................... 46
         :::::::::|||   ::::::||| 
     356 TATCAAAAACATGATTACAGGGACATCTCAGGCTGACTGTGCTGTCCTGA 405

      47 .................ThrAsnLysGlnGluIleSerSerArgPheGl 57
                          :::|||   :::   :::|||   |||:::|| 
     406 TTGTTGCTGCTGGTGTTGGTGAATTTGAAGCTGGTATCTCCAAGAATGGG 455

      57 ySerAlaAspSerIleSerPheMetIle..................... 66
         |      :::|||:::   |||   ::: 
     456 CAGACCCGAGAGCATGCCCTTCTGGCTTACACACTGGGTGTGAAACAACT 505

      67 ........ValValIleLysPhe.GlyHisThrAlaIleLeuAlaPro.A 80
                   |||:::   |||:::    |||   |||   |||   ||| : 
     506 AATTGTCGGTGTTAACAAAATGGATTCCACTGAGCCACCCTACAGCCAGA 555

      80 snArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrpValLys 96
         ::|||:::::::|||||::::::   :::   ::::::::: 
     556 AGAGATATGAGGAAATTGTTAAGGAAGTCAGCACTTACATTAAGAAAATT 605

      97 ProTyrArgProLysAsnLeuSerPhe 105
            |||:::|||   :::::::::||| 
     606 GGCTACAACCCCGACACAGTAGCATTT 632

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-484-101B-49

seq_documentation_block:
; Sequence 49, Application US/08484101B
; Patent No. 5824868
;  GENERAL INFORMATION:
;    APPLICANT:  California Institute of Technology
;    TITLE OF INVENTION:  PLANTS HAVING MODIFIED RESPONSE TO
;    TITLE OF INVENTION:  ETHYLENE
;    NUMBER OF SEQUENCES:  50
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard F. Trecartin
;      STREET:  3400 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,101B
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/
;      FILING DATE:  01-JUL-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/086,555
;      FILING DATE:  01-JUL-1993
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Trecartin, Richard F.
;      REGISTRATION NUMBER:  31,801
;      REFERENCE/DOCKET NUMBER:  A-57515-2/RFT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;  INFORMATION FOR SEQ ID NO:  49:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2405 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  288..2196
US-08-484-101B-49


alignment_scores:
            Quality:   59.50             Length:     84
              Ratio:  1.526               Gaps:      2
 Percent Similarity:  46.429   Percent Identity: 23.810

alignment_block:
 US-10-048-197-2 x US-08-484-101B-49   ..

 Align seg 1/1  to: US-08-484-101B-49  from: 1  to: 2405

      22 CysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThr.. 37
         ||||||::::::||||||      :::||||||   ::::::::::|||
     195 TGTGTCGCAGCAGGAAATAGCATCTTAAAGGAAGGAAGGAAGGAAACTCG 244

      37 ..................................................... 37

     245 AAAGTTACTAAAAATTTTTGATTCTTTGGGACGAAACGAGATAATGGAAT 294

      38 ..........IleGluGlnThrIleIleLysGlyLys......ThrAsn 48
                   ||||||      :::      |||         ::::::
     295 CCTGTGATTGCATTGAGGCTTTACTGCCAACTGGTGACCTGCTGGTTAAA 344

      49 LysGlnGluIleSerSerArgPheGlySerAlaAspSerIleSerPheMe 65
            |||   :::|||:::   |||   ::::::      :::|||:::::
     345 TACCAATACCTCTCAGATTTCTTCATTGCTGTAGCCTACTTTTCCATTCT 394

      65 tIleValValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgT 82
         ::::   :::|||   |||   |||   :::          |||   ||||
     395 GTTGGAGCTTATTTATTTTGTCCACAAATCTGCATGCTTCCCATACAGAT 444

      82 rp 82
         ||
     445 GG 446

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-714-524D-49

seq_documentation_block:
; Sequence 49, Application US/08714524D
; Patent No. 6294716
; GENERAL INFORMATION:
;  APPLICANT: Meyerowitz, Elliott M
;  APPLICANT:  Chang, Caren
;  APPLICANT:  Bleecker, Anthony B
;  TITLE OF INVENTION: PLANTS HAVING MODIFIED RESPONSE TO ETHYLENE
;  FILE REFERENCE: a-57515-4
;  CURRENT APPLICATION NUMBER: US/08/714,524D
;  CURRENT FILING DATE:  1996-09-16
;  NUMBER OF SEQ ID NOS: 56
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
;   LENGTH: 2405
;   TYPE: DNA
;   ORGANISM: Lycopersicon esculentum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (288)..(2195)
US-08-714-524D-49


alignment_scores:
            Quality:   59.50             Length:     84
              Ratio:  1.526               Gaps:      2
 Percent Similarity:  46.429   Percent Identity: 23.810

alignment_block:
 US-10-048-197-2 x US-08-714-524D-49   ..

 Align seg 1/1  to: US-08-714-524D-49  from: 1  to: 2405

      22 CysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThr.. 37
         ||||||::::::||||||      :::||||||   ::::::::::|||
     195 TGTGTCGCAGCAGGAAATAGCATCTTAAAGGAAGGAAGGAAGGAAACTCG 244

      37 ..................................................... 37

     245 AAAGTTACTAAAAATTTTTGATTCTTTGGGACGAAACGAGATAATGGAAT 294

      38 ..........IleGluGlnThrIleIleLysGlyLys......ThrAsn 48
                   ||||||      :::      |||         ::::::
     295 CCTGTGATTGCATTGAGGCTTTACTGCCAACTGGTGACCTGCTGGTTAAA 344

      49 LysGlnGluIleSerSerArgPheGlySerAlaAspSerIleSerPheMe 65
            |||   :::|||:::   |||   ::::::      :::|||:::::
     345 TACCAATACCTCTCAGATTTCTTCATTGCTGTAGCCTACTTTTCCATTCT 394

      65 tIleValValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgT 82
         ::::   :::|||   |||   |||   :::          |||   ||||
     395 GTTGGAGCTTATTTATTTTGTCCACAAATCTGCATGCTTCCCATACAGAT 444

      82 rp 82
         ||
     445 GG 446

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-058-489-19

seq_documentation_block:
; Sequence 19, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
;  APPLICANT: Whitehead Institute for Biomedical Research
;  APPLICANT: Lahn, Bruce
;  APPLICANT: Page, David
;  TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
;  TITLE OF INVENTION: the Y Chromosome
;  FILE REFERENCE: WHI97-08pA
;  CURRENT APPLICATION NUMBER: US/09/058,489
;  CURRENT FILING DATE: 1998-04-10
;  EARLIER APPLICATION NUMBER: 60/041,877
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 91
```

```
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19
;   LENGTH: 4335
;   TYPE: DNA
;   ORGANISM: Human
US-09-058-489-19


alignment_scores:
            Quality:   59.50              Length:    100
              Ratio:  1.044                Gaps:      3
 Percent Similarity:  57.000   Percent Identity: 26.000

alignment_block:
 US-10-048-197-2 x US-09-058-489-19   ..

 Align seg 1/1  to: US-09-058-489-19  from: 1  to: 4335

       9 ThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerTh 25
         |||   :::   ||||||   :::      ::::::|||      |||::
    2614 ACAGACACTATCTTGCTAGGCAGTAATTGTATAGCAGGAAGTGAAAGTAA 2663

      25 rGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIle......GluG 40
         :|||||||||         :::||||||   :::|||:::      ::::
    2664 TGGAAATGTGCCTTACCTGCAGCAAAATACACACACTCTACCTCATAATC 2713

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
         ::|||   ::::::::::::|||::::::       ::::::
    2714 ATACAGACCTGAACAGCAGCACAGAAGAGCCATGGAGAAAACAGCTATCT 2763

      57 GlySerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHi 73
         :::|||||||:::::::::                  |||
    2764 AACTCCGCTCAGGGGCTTCAT.................AAAAGTCAGAG 2795

      73 sThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleI 90
          :::::::::   :::||||||:::   |||   ::::::|||
    2796 TTCATGTTTGTCAGGACCTAATGAAGAACAACCTCTGTTTTCCAC..... 2840

      90 leSerPheLeuTrpValLysProTyrArgProLysAsnLeuSerPheTyr 106
                   |||||||:::|||      |||   |||      :::|||
    2841 ...........TGGGTCAGCCCAGTATCACCAGGCAACTAGCACTGGTAT 2879

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-058-489-20

seq_documentation_block:
; Sequence 20, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
;  APPLICANT: Whitehead Institute for Biomedical Research
;  APPLICANT: Lahn, Bruce
;  APPLICANT: Page, David
;  TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
;  TITLE OF INVENTION: the Y Chromosome
;  FILE REFERENCE: WHI97-08pA
;  CURRENT APPLICATION NUMBER: US/09/058,489
;  CURRENT FILING DATE: 1998-04-10
;  EARLIER APPLICATION NUMBER: 60/041,877
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 91
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 20
;   LENGTH: 4931
;   TYPE: DNA
;   ORGANISM: Human
US-09-058-489-20


alignment_scores:
            Quality:   59.50              Length:    100
              Ratio:  1.044                Gaps:      3
 Percent Similarity:  57.000   Percent Identity: 26.000

alignment_block:
 US-10-048-197-2 x US-09-058-489-20   ..

 Align seg 1/1  to: US-09-058-489-20  from: 1  to: 4931

       9 ThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerTh 25
         |||   :::   ||||||   :::      ::::::|||      |||::
    2614 ACAGACACTATCTTGCTAGGCAGTAATTGTATAGCAGGAAGTGAAAGTAA 2663

      25 rGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIle......GluG 40
         :|||||||||         :::||||||   :::|||:::      ::::
    2664 TGGAAATGTGCCTTACCTGCAGCAAAATACACACACTCTACCTCATAATC 2713

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
         ::|||   ::::::::::::|||::::::       ::::::
    2714 ATACAGACCTGAACAGCAGCACAGAAGAGCCATGGAGAAAACAGCTATCT 2763

      57 GlySerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHi 73
         :::|||||||:::::::::                  |||
    2764 AACTCCGCTCAGGGGCTTCAT.................AAAAGTCAGAG 2795

      73 sThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleI 90
          :::::::::   :::||||||:::   |||   ::::::|||
    2796 TTCATGTTTGTCAGGACCTAATGAAGAACAACCTCTGTTTTCCAC..... 2840

      90 leSerPheLeuTrpValLysProTyrArgProLysAsnLeuSerPheTyr 106
                   |||||||:::|||      |||   |||      :::|||
    2841 ...........TGGGTCAGCCCAGTATCACCAGGCAACTAGCACTGGTAT 2879

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-058-489-21

seq_documentation_block:
; Sequence 21, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
;  APPLICANT: Whitehead Institute for Biomedical Research
;  APPLICANT: Lahn, Bruce
;  APPLICANT: Page, David
;  TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
;  TITLE OF INVENTION: the Y Chromosome
;  FILE REFERENCE: WHI97-08pA
;  CURRENT APPLICATION NUMBER: US/09/058,489
;  CURRENT FILING DATE: 1998-04-10
;  EARLIER APPLICATION NUMBER: 60/041,877
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 91
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 21
;   LENGTH: 6476
;   TYPE: DNA
;   ORGANISM: Human
US-09-058-489-21


alignment_scores:
            Quality:   59.50              Length:    100
              Ratio:  1.044                Gaps:      3
 Percent Similarity:  57.000   Percent Identity: 26.000

alignment_block:
 US-10-048-197-2 x US-09-058-489-21   ..

 Align seg 1/1  to: US-09-058-489-21  from: 1  to: 6476

       9 ThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerTh 25
         |||   :::   ||||||   :::      ::::::|||      |||::
    2614 ACAGACACTATCTTGCTAGGCAGTAATTGTATAGCAGGAAGTGAAAGTAA 2663

      25 rGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIle......GluG 40
         :|||||||||         :::||||||   :::|||:::      ::::
    2664 TGGAAATGTGCCTTACCTGCAGCAAAATACACACACTCTACCTCATAATC 2713

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
         ::|||   ::::::::::::|||::::::       ::::::
```

```
    2714 ATACAGACCTGAACAGCAGCACAGAAGAGCCATGGAGAAAACAGCTATCT 2763

      57 GlySerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHi 73
         :::||||||:::::::::                  |||
    2764 AACTCCGCTCAGGGGCTTCAT.................AAAAGTCAGAG 2795

      73 sThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleI 90
         :::::::::   :::|||||:::   |||   :::::|||
    2796 TTCATGTTTGTCAGGACCTAATGAAGAACAACCTCTGTTTTCCAC..... 2840

      90 leSerPheLeuTrpValLysProTyrArgProLysAsnLeuSerPheTyr 106
                 ||||||:::|||      |||   |||      :::|||
    2841 ..........TGGGTCAGCCCAGTATCACCAGGCAACTAGCACTGGTAT 2879

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-961-522-3

seq_documentation_block:
; Sequence 3, Application US/07961522
; Patent No. 5417971
;  GENERAL INFORMATION:
;    APPLICANT:  POTTER, ANDREW A.
;    APPLICANT:  GERLACH, GERALD F.
;    APPLICANT:  WILLSON, PHILIP J.
;    APPLICANT:  ROSSI-CAMPOS, AMALIA
;    TITLE OF INVENTION:  VACCINES FOR ACTINOBACILLUS
;    TITLE OF INVENTION:  PLEUROPNEUMONIAE
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ROBERTA L. ROBINS
;      STREET:  635 BRYANT STREET
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/961,522
;      FILING DATE:  19921015
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ROBINS, ROBERTA L.
;      REGISTRATION NUMBER:  33,208
;      REFERENCE/DOCKET NUMBER:  9000-0015.20
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 617-8999
;      TELEFAX:  (415) 327-3231
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1903 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1777
US-07-961-522-3


alignment_scores:
            Quality:   59.00            Length:     76
              Ratio:  1.439              Gaps:      4
 Percent Similarity:  53.947   Percent Identity: 27.632

alignment_block:
 US-10-048-197-2 x US-07-961-522-3   ..

 Align seg 1/1  to: US-07-961-522-3  from: 1  to: 1903
```

```
      38 IleGluGlnThrIleIle......LysGlyLysThrAsnLysGlnGluIl 52
         ::::::   |||::::::      |||   |||   :::||||||   ||
     878 TTGCAGGCAACCGTTTTAGGGGAAAAGCCAAAGCAGAAAAAGCAGGTGAT 927

      52 eSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValValI 69
         |                          |||::::::|||   :::|
     928 CCGA........................TCTTTACTGATTCAAATTA 950

      69 leLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln...... 83
         ||      |||   :::   :::|||      ::::::||||||
     951 TCTTGAAGGGGGATTCTATGGTCCTAAAGCTGAAGAAATGGCAGGGAAGT 1000

      84 ......GluIleLeuSerLeuIleIleSerPheLeuTrpValLys..... 96
               :::|||::::::|||   ::::::::::|||   :::|||
    1001 TTTTCACAAATAATAAATCTCTCTTTGCAGTATTTGCAGCTAAAAGTGAA 1050

      97 ....ProTyrArgProLysAsnLeuSer 104
                 ||||||:::|||   |||
    1051 AACGGCGAGACGACCACAGAACGAATCA 1078

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-217-438-3

seq_documentation_block:
; Sequence 3, Application US/08217438
; Patent No. 5521072
;  GENERAL INFORMATION:
;    APPLICANT:  Potter, Andrew A.
;    APPLICANT:  Gerlach, Gerald F.
;    APPLICANT:  Willson, Philp J.
;    APPLICANT:  Rossi-Campos, Amalia
;    TITLE OF INVENTION:  ACTINOBACILLUS PLEUROPNEIMONIAE
;    TITLE OF INVENTION:  TRANSFERRIN BINDING PROTEINS AND USES THEREOF
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Reed & Robins
;      STREET:  635 Bryant Street
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/217,438
;      FILING DATE:  22-MAR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Robins, Roberta L.
;      REGISTRATION NUMBER:  33,208
;      REFERENCE/DOCKET NUMBER:  9001-0015.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-617-8999
;      TELEFAX:  415-327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1903 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1777
US-08-217-438-3


alignment_scores:
            Quality:   59.00            Length:     76
              Ratio:  1.439              Gaps:      4
 Percent Similarity:  53.947   Percent Identity: 27.632
```

```
alignment_block:
 US-10-048-197-2 x US-08-217-438-3   ..

 Align seg 1/1  to: US-08-217-438-3  from: 1  to: 1903

      38 IleGluGlnThrIleIle......LysGlyLysThrAsnLysGlnGluIl 52
         ::::::   |||::::::      |||   |||   :::||||||   ||
     878 TTGCAGGCAACCGTTTTAGGGGAAAAGCCAAAGCAGAAAAAGCAGGTGAT 927

      52 eSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValValI 69
         |                              |||:::::|||   :::|
     928 CCGA.........................TCTTTACTGATTCAAATTA 950

      69 leLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln...... 83
         ||      |||   :::   :::|||      ::::::||||||
     951 TCTTGAAGGGGGATTCTATGGTCCTAAAGCTGAAGAAATGGCAGGGAAGT 1000

      84 ......GluIleLeuSerLeuIleIleSerPheLeuTrpValLys..... 96
              :::|||::::::|||   :::::::::|||   :::|||
    1001 TTTTCACAAATAATAAATCTCTCTTTGCAGTATTTGCAGCTAAAAGTGAA 1050

      97 ....ProTyrArgProLysAsnLeuSer 104
                 ||||||:::|||   |||
    1051 AACGGCGAGACGACCACAGAACGAATCA 1078

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-321-978-3

seq_documentation_block:
; Sequence 3, Application US/08321978
; Patent No. 5801018
;  GENERAL INFORMATION:
;    APPLICANT:  POTTER, ANDREW A.
;    APPLICANT:  GERLACH, GERALD F.
;    APPLICANT:  WILLSON, PHILIP J.
;    APPLICANT:  ROSSI-CAMPOS, AMALIA
;    TITLE OF INVENTION:  VACCINES FOR ACTINOBACILLUS
;    TITLE OF INVENTION:  PLEUROPNEUMONIAE
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ROBERTA L. ROBINS
;      STREET:  285 HAMILTON AVE, SUITE 200
;      CITY:  PALO ALTO.
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/321,978
;      FILING DATE:  12-OCT-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ROBINS, ROBERTA L.
;      REGISTRATION NUMBER:  33,208
;      REFERENCE/DOCKET NUMBER:  9001-0015.02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 327-3400
;      TELEFAX:  (415) 327-3231
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1903 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1777
US-08-321-978-3

alignment_scores:
            Quality:   59.00            Length:     76
              Ratio:  1.439              Gaps:      4
 Percent Similarity:  53.947   Percent Identity: 27.632

alignment_block:
 US-10-048-197-2 x US-08-321-978-3   ..

 Align seg 1/1  to: US-08-321-978-3  from: 1  to: 1903

      38 IleGluGlnThrIleIle......LysGlyLysThrAsnLysGlnGluIl 52
         ::::::   |||::::::      |||   |||   :::||||||   ||
     878 TTGCAGGCAACCGTTTTAGGGGAAAAGCCAAAGCAGAAAAAGCAGGTGAT 927

      52 eSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValValI 69
         |                              |||:::::|||   :::|
     928 CCGA.........................TCTTTACTGATTCAAATTA 950

      69 leLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln...... 83
         ||      |||   :::   :::|||      ::::::||||||
     951 TCTTGAAGGGGGATTCTATGGTCCTAAAGCTGAAGAAATGGCAGGGAAGT 1000

      84 ......GluIleLeuSerLeuIleIleSerPheLeuTrpValLys..... 96
              :::|||::::::|||   :::::::::|||   :::|||
    1001 TTTTCACAAATAATAAATCTCTCTTTGCAGTATTTGCAGCTAAAAGTGAA 1050

      97 ....ProTyrArgProLysAsnLeuSer 104
                 ||||||:::|||   |||
    1051 AACGGCGAGACGACCACAGAACGAATCA 1078

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-710-584-3

seq_documentation_block:
; Sequence 3, Application US/08710584
; Patent No. 5876725
;  GENERAL INFORMATION:
;    APPLICANT:  POTTER, ANDREW A.
;    APPLICANT:  GERLACH, GERALD F.
;    APPLICANT:  WILLSON, PHILIP J.
;    APPLICANT:  ROSSI-CAMPOS, AMALIA
;    TITLE OF INVENTION:  VACCINES FOR ACTINOBACILLUS
;    TITLE OF INVENTION:  PLEUROPNEUMONIAE
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ROBERTA L. ROBINS
;      STREET:  285 HAMILTON AVE, SUITE 200
;      CITY:  PALO ALTO
;      STATE:  CALIFORNIA
;      COUNTRY:  UNITED STATES OF AMERICA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/710,584
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/321,978
;      FILING DATE:  12-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ROBINS, ROBERTA L.
;      REGISTRATION NUMBER:  33,208
;      REFERENCE/DOCKET NUMBER:  9001-0015.02
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 327-3400
;      TELEFAX:  (415) 327-3231
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
```

```
;       LENGTH:  1903 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..1777
US-08-710-584-3


alignment_scores:
            Quality:   59.00              Length:     76
              Ratio:  1.439                Gaps:      4
 Percent Similarity:  53.947   Percent Identity: 27.632

alignment_block:
 US-10-048-197-2 x US-08-710-584-3    ..

 Align seg 1/1  to: US-08-710-584-3  from: 1  to: 1903

      38 IleGluGlnThrIleIle......LysGlyLysThrAsnLysGlnGluIl 52
         ::::::   |||::::::      |||   |||   :::||||||   ||
     878 TTGCAGGCAACCGTTTTAGGGGAAAAGCCAAAGCAGAAAAAGCAGGTGAT 927

      52 eSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValValI 69
         |                          |||:::::::|||   :::|
     928 CCGA.........................TCTTTACTGATTCAAATTA 950

      69 leLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln...... 83
         ||       |||   :::   :::|||      ::::::|||||
     951 TCTTGAAGGGGGATTCTATGGTCCTAAAGCTGAAGAAATGGCAGGGAAGT 1000

      84 ......GluIleLeuSerLeuIleIleSerPheLeuTrpValLys..... 96
            :::|||::::::|||   ::::::::::|||   :::|||
    1001 TTTTCACAAATAATAAATCTCTCTTTGCAGTATTTGCAGCTAAAAGTGAA 1050

      97 ....ProTyrArgProLysAsnLeuSer 104
                |||||:::|||   |||
    1051 AACGGCGAGACGACCACAGAACGAATCA 1078

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-564-805-223

seq_documentation_block:
; Sequence 223, Application US/09564805
; Patent No. 6333403
; GENERAL INFORMATION:
;  APPLICANT:  Tavtigian, Sean V.
;  APPLICANT:  Teng, David H.F.
;  APPLICANT:  Simard, Jacques
;  APPLICANT:  Rommens, Johanna M.
;  APPLICANT:  Myriad Genetics, Inc.
;  TITLE OF INVENTION:  Chromosome 17p-Linked Prostate Cancer Susceptibility
;  TITLE OF INVENTION:  Gene and a Paralog and Orthologous Genes
;  FILE REFERENCE:  2318-258
;  CURRENT APPLICATION NUMBER: US/09/564,805
;  CURRENT FILING DATE:  2000-05-05
;  PRIOR APPLICATION NUMBER:  US 60/107,468
;  PRIOR FILING DATE:  1998-11-06
;  PRIOR APPLICATION NUMBER:  09/434,382
;  PRIOR FILING DATE:  1999-11-05
;  NUMBER OF SEQ ID NOS:  240
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 223
;   LENGTH: 2908
;   TYPE: DNA
;   ORGANISM: Pan troglodytes
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(2478)
US-09-564-805-223


alignment_scores:
            Quality:   59.00              Length:     73
              Ratio:  1.686                Gaps:      1
 Percent Similarity:  47.945   Percent Identity: 26.027

alignment_block:
 US-10-048-197-2 x US-09-564-805-223    ..

 Align seg 1/1  to: US-09-564-805-223  from: 1  to: 2908

      22 CysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIl 38
         |||   |||||||||:::|||            |||   ||||||
    1379 TGCAGGAGTACAGGAGGAGTGCGCAGGACGGCCCAGCCCCAGCAGAGAAA 1428

      38 eGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerA 55
          |||   :::      |||:::::::::   :::|||   :::   ||||
    1429 AGAAGTCAGTACCCAGAAATCATCTTCCTTGGAACAGGGTCTGCCATCCC 1478

      55 rgPheGlySerAlaAspSerIleSerPheMet..................... 65
         ||               |||:::     :::
    1479 GATGAAGATTCGAAATGTCAGTGCCACACTTGTCAACATAAGCCCCGACA 1528

      66 ..................IleValValIleLysPheGlyHisThrAlaIl 76
                           :::|||      ::::::|||      |||::
    1529 CGTCTCTGCTACTGGACTGTGGTGAGGGCACGTTTGGGCAGCTGTGCCGT 1578

      76 eLeuAlaProAsnArgTrp 82
         :::::::   :::||||||
    1579 CATTACGGAGACCAGGTGG 1597

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-121-713D-59

seq_documentation_block:
; Sequence 59, Application US/08121713D
; Patent No. 5639856
;  GENERAL INFORMATION:
;    APPLICANT:  Goodman, Corey S.
;    APPLICANT:  Kolodkin, Alex L.
;    APPLICANT:  Matthes, David
;    APPLICANT:  Bentley, David R.
;    APPLICANT:  O'Connor, Timothy
;    TITLE OF INVENTION:  The Semaphorin Gene Family
;    NUMBER OF SEQUENCES:  100
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/121,713D
;      FILING DATE:  13-SEP-1993
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A.
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  B94-002-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415) 343-4342
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
```

```
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1953
US-08-121-713D-59


alignment_scores:
           Quality:   59.00              Length:    104
             Ratio:  1.035                Gaps:      6
 Percent Similarity:  54.808   Percent Identity: 26.923

alignment_block:
 US-10-048-197-2 x US-08-121-713D-59/rev   ..

 Align seg 1/1  to reverse of: US-08-121-713D-59  from: 1  to: 3560

       6 IleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCy 22
         :::|||:::   |||      ::::::      :::||||||||||
    1650 CTCCGTATCCGGATACGGCAGATTATCATCCTCGTCCTTGTGACA..... 1606

      22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleG 39
         ::::::|||:::::::|||||   ||||||   :::|||||||      :
    1605 .ACGGCGACCGCAGAAGTAGCCTGTAAAGAAGCCGACCAGCAGCGAAAAG 1557

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGlu.IleSerSerAr 55
         :::::|||         |||:::::::::   :::||| :::|||:::
    1556 ATCGAACC.........GGCCAGAACGGCCATCACGAGGGTCTCCACAGT 1516

      55 gPheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheG 72
         :::
    1515 GTAC..................................................... 1512

      72 lyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeu 88
               |||::::::      :::::::::   :::||||||||||:::
    1511 .....TGGGCGTTGATTATATCTTCAAAGTTCTTATCAATATTGTCGATT 1467

      89 IleIle...SerPheLeuTrpValLysProTyrArgProLysAsnLeuSe 104
         |||      |||:::||||||            |||   :::|||   ||
    1466 ATTTCCTGATCCTTGCTCTGG.........CGTCGCGAATCCAATAAGTC 1426

     104 rPheTyrLeu 107
         |:::   |||
    1425 CATGTCGTTG 1416

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-835-268-59

seq_documentation_block:
; Sequence 59, Application US/08835268
; Patent No. 5807826
;  GENERAL INFORMATION:
;    APPLICANT:  Goodman, Corey S.
;    APPLICANT:  Kolodkin, Alex L.
;    APPLICANT:  Matthes, David
;    APPLICANT:  Bentley, David R.
;    APPLICANT:  O'Connor, Timothy
;    TITLE OF INVENTION:  The Semaphorin Gene Family
;    NUMBER OF SEQUENCES:  100
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/835,268
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/121,713
;      FILING DATE:  13-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A.
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  B94-002-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415) 343-4342
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1953
US-08-835-268-59


alignment_scores:
           Quality:   59.00              Length:    104
             Ratio:  1.035                Gaps:      6
 Percent Similarity:  54.808   Percent Identity: 26.923

alignment_block:
 US-10-048-197-2 x US-08-835-268-59/rev   ..

 Align seg 1/1  to reverse of: US-08-835-268-59  from: 1  to: 3560

       6 IleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCy 22
         :::|||:::   |||      ::::::      :::||||||||||
    1650 CTCCGTATCCGGATACGGCAGATTATCATCCTCGTCCTTGTGACA..... 1606

      22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleG 39
         ::::::|||:::::::|||||   ||||||   :::|||||||      :
    1605 .ACGGCGACCGCAGAAGTAGCCTGTAAAGAAGCCGACCAGCAGCGAAAAG 1557

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGlu.IleSerSerAr 55
         :::::|||         |||:::::::::   :::||| :::|||:::
    1556 ATCGAACC.........GGCCAGAACGGCCATCACGAGGGTCTCCACAGT 1516

      55 gPheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheG 72
         :::
    1515 GTAC..................................................... 1512

      72 lyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeu 88
               |||::::::      :::::::::   :::||||||||||:::
    1511 .....TGGGCGTTGATTATATCTTCAAAGTTCTTATCAATATTGTCGATT 1467

      89 IleIle...SerPheLeuTrpValLysProTyrArgProLysAsnLeuSe 104
         |||      |||:::||||||            |||   :::|||   ||
    1466 ATTTCCTGATCCTTGCTCTGG.........CGTCGCGAATCCAATAAGTC 1426

     104 rPheTyrLeu 107
         |:::   |||
    1425 CATGTCGTTG 1416

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-060-692-59

seq_documentation_block:
; Sequence 59, Application US/09060692
; Patent No. 5935865
;  GENERAL INFORMATION:
;    APPLICANT:  Goodman, Corey S.
;    APPLICANT:  Kolodkin, Alex L.
;    APPLICANT:  Matthes, David
;    APPLICANT:  Bentley, David R.
```

```
;    APPLICANT:  O'Connor, Timothy
;    TITLE OF INVENTION:  The Semaphorin Gene Family
;    NUMBER OF SEQUENCES:  100
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/060,692
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/121,713
;      FILING DATE:  13-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A.
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  B94-002-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415) 343-4342
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1953
US-09-060-692-59


alignment_scores:
           Quality:   59.00              Length:    104
             Ratio:  1.035                Gaps:      6
 Percent Similarity:  54.808   Percent Identity: 26.923

alignment_block:
 US-10-048-197-2 x US-09-060-692-59/rev   ..

 Align seg 1/1  to reverse of: US-09-060-692-59  from: 1  to: 3560

       6 IleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCy 22
         :::|||:::   |||      ::::::      :::||||||||||
    1650 CTCCGTATCCGGATACGGCAGATTATCATCCTCGTCCTTGTGACA..... 1606

      22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleG 39
         ::::::|||::::::|||||   ||||||   :::||||||      :
    1605 .ACGGCGACCGCAGAAGTAGCCTGTAAAGAAGCCGACCAGCAGCGAAAAG 1557

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGlu.IleSerSerAr 55
         :::::|||         |||:::::::::   :::||| :::|||:::
    1556 ATCGAACC.........GGCCAGAACGGCCATCACGAGGGTCTCCACAGT 1516

      55 gPheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheG 72
         :::
    1515 GTAC..................................................... 1512

      72 lyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeu 88
                |||::::::      :::::::::   :::|||||||||:::
    1511 .....TGGGCGTTGATTATATCTTCAAAGTTCTTATCAATATTGTCGATT 1467

      89 IleIle...SerPheLeuTrpValLysProTyrArgProLysAsnLeuSe 104
         |||      |||:::||||||            |||   :::|||   ||
    1466 ATTTCCTGATCCTTGCTCTGG.........CGTCGCGAATCCAATAAGTC 1426

     104 rPheTyrLeu 107
         |:::   |||
    1425 CATGTCGTTG 1416

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-833-391-59

seq_documentation_block:
; Sequence 59, Application US/08833391
; Patent No. 6013781
;  GENERAL INFORMATION:
;    APPLICANT:  Goodman, Corey S.
;    APPLICANT:  Kolodkin, Alex L.
;    APPLICANT:  Matthes, David
;    APPLICANT:  Bentley, David R.
;    APPLICANT:  O'Connor, Timothy
;    TITLE OF INVENTION:  The Semaphorin Gene Family
;    NUMBER OF SEQUENCES:  100
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/833,391
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/121,713
;      FILING DATE:  13-SEP-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A.
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  B94-002-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415) 343-4342
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1953
US-08-833-391-59


alignment_scores:
           Quality:   59.00              Length:    104
             Ratio:  1.035                Gaps:      6
 Percent Similarity:  54.808   Percent Identity: 26.923

alignment_block:
 US-10-048-197-2 x US-08-833-391-59/rev   ..

 Align seg 1/1  to reverse of: US-08-833-391-59  from: 1  to: 3560

       6 IleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCy 22
         :::|||:::   |||      ::::::      :::||||||||||
```

```
    1650 CTCCGTATCCGGATACGGCAGATTATCATCCTCGTCCTTGTGACA..... 1606

      22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleG 39
         :::::::|||:::::::||||||   ||||||   :::||||||      :
    1605 .ACGGCGACCGCAGAAGTAGCCTGTAAAGAAGCCGACCAGCAGCGAAAAG 1557

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGlu.IleSerSerAr 55
         :::::|||         |||:::::::::   :::||| :::|||:::
    1556 ATCGAACC.........GGCCAGAACGGCCATCACGAGGGTCTCCACAGT 1516

      55 gPheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheG 72
         :::
    1515 GTAC..................................................... 1512

      72 lyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeu 88
               |||::::::      :::::::::   :::||||||||||:::
    1511 .....TGGGCGTTGATTATATCTTCAAAGTTCTTATCAATATTGTCGATT 1467

      89 IleIle...SerPheLeuTrpValLysProTyrArgProLysAsnLeuSe 104
         |||      |||:::||||||            |||   :::|||   ||
    1466 ATTTCCTGATCCTTGCTCTGG.........CGTCGCGAATCCAATAAGTC 1426

     104 rPheTyrLeu 107
         |:::   |||
    1425 CATGTCGTTG 1416

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-060-610-59

seq_documentation_block:
; Sequence 59, Application US/09060610
; Patent No. 6344544
;  GENERAL INFORMATION:
;    APPLICANT:  Goodman, Corey S.
;    APPLICANT:  Kolodkin, Alex L.
;    APPLICANT:  Matthes, David
;    APPLICANT:  Bentley, David R.
;    APPLICANT:  O'Connor, Timothy
;    TITLE OF INVENTION:  The Semaphorin Gene Family
;    NUMBER OF SEQUENCES:  100
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  268 Bush Street, Suite 3200
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/060,610
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/835,268
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A.
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  B94-002-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)343-4341
;      TELEFAX:  (415) 343-4342
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1953
US-09-060-610-59


alignment_scores:
            Quality:   59.00            Length:    104
              Ratio:  1.035              Gaps:      6
 Percent Similarity:  54.808   Percent Identity: 26.923

alignment_block:
 US-10-048-197-2 x US-09-060-610-59/rev   ..

 Align seg 1/1  to reverse of: US-09-060-610-59  from: 1  to: 3560

       6 IleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCy 22
         :::|||:::   |||      ::::::      :::||||||||||
    1650 CTCCGTATCCGGATACGGCAGATTATCATCCTCGTCCTTGTGACA..... 1606

      22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleG 39
         :::::::|||:::::::||||||   ||||||   :::||||||      :
    1605 .ACGGCGACCGCAGAAGTAGCCTGTAAAGAAGCCGACCAGCAGCGAAAAG 1557

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGlu.IleSerSerAr 55
         :::::|||         |||:::::::::   :::||| :::|||:::
    1556 ATCGAACC.........GGCCAGAACGGCCATCACGAGGGTCTCCACAGT 1516

      55 gPheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheG 72
         :::
    1515 GTAC..................................................... 1512

      72 lyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeu 88
               |||::::::      :::::::::   :::||||||||||:::
    1511 .....TGGGCGTTGATTATATCTTCAAAGTTCTTATCAATATTGTCGATT 1467

      89 IleIle...SerPheLeuTrpValLysProTyrArgProLysAsnLeuSe 104
         |||      |||:::||||||            |||   :::|||   ||
    1466 ATTTCCTGATCCTTGCTCTGG.........CGTCGCGAATCCAATAAGTC 1426

     104 rPheTyrLeu 107
         |:::   |||
    1425 CATGTCGTTG 1416

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US94-10151A-59

seq_documentation_block:
; Sequence 59, Application PC/TUS9410151A
;  GENERAL INFORMATION:
;    APPLICANT:  The Regents of the University of California
;    TITLE OF INVENTION:  The Semaphorin Gene Family
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR HOHBACH TEST ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/10151A
;      FILING DATE:  13-SEP-1994
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard A.
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  FP-58750-PC/RAO
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299 FHT UR
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1953
PCT-US94-10151A-59


alignment_scores:
            Quality:   59.00              Length:    104
              Ratio:  1.035                Gaps:      6
 Percent Similarity:  54.808   Percent Identity: 26.923

alignment_block:
 US-10-048-197-2 x PCT-US94-10151A-59/rev   ..

 Align seg 1/1  to reverse of: PCT-US94-10151A-59  from: 1  to: 3560

       6 IleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCy 22
         :::|||:::   |||      :::::      :::||||||||||
    1650 CTCCGTATCCGGATACGGCAGATTATCATCCTCGTCCTTGTGACA..... 1606

      22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleG 39
         ::::::|||:::::|||||   ||||||   :::||||||      :
    1605 .ACGGCGACCGCAGAAGTAGCCTGTAAAGAAGCCGACCAGCAGCGAAAAG 1557

      39 luGlnThrIleIleLysGlyLysThrAsnLysGlnGlu.IleSerSerAr 55
         :::::|||         |||:::::::::   :::||| :::|||:::
    1556 ATCGAACC.........GGCCAGAACGGCCATCACGAGGGTCTCCACAGT 1516

      55 gPheGlySerAlaAspSerIleSerPheMetIleValValIleLysPheG 72
         :::
    1515 GTAC..................................................... 1512

      72 lyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeu 88
               |||::::::      :::::::::   :::||||||||||:::
    1511 .....TGGGCGTTGATTATATCTTCAAAGTTCTTATCAATATTGTCGATT 1467

      89 IleIle...SerPheLeuTrpValLysProTyrArgProLysAsnLeuSe 104
         |||      |||:::||||||            |||   :::|||   ||
    1466 ATTTCCTGATCCTTGCTCTGG.........CGTCGCGAATCCAATAAGTC 1426

     104 rPheTyrLeu 107
         |:::   |||
    1425 CATGTCGTTG 1416

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-007-999-1

seq_documentation_block:
; Sequence 1, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
;  APPLICANT: Nichols, Scott E.
;  TITLE OF INVENTION: Substitutes for Modified Starch and
;  TITLE OF INVENTION: Latexes in Paper Manufacture
;  FILE REFERENCE: 0356D
;  CURRENT APPLICATION NUMBER: US/09/007,999
;  CURRENT FILING DATE: 1998-01-16
;  EARLIER APPLICATION NUMBER: 08/478,704
;  EARLIER FILING DATE: 1995-06-07
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 4460
;   TYPE: DNA
;   ORGANISM: Streptococcus mutans
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (9)...(4434)
US-09-007-999-1


alignment_scores:
            Quality:   59.00              Length:     49
              Ratio:  2.034                Gaps:      2
 Percent Similarity:  59.184   Percent Identity: 32.653

alignment_block:
 US-10-048-197-2 x US-09-007-999-1   ..

 Align seg 1/1  to: US-09-007-999-1  from: 1  to: 4460

       9 ThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerTh 25
         |||:::   :::||||||   :::      |||:::||||||:::|||
    3643 ACCGTAATCGTTTTATTGAAAACGAAGAAGGTAAATGGCTGTATCTCG.. 3690

      25 rGlyAsnValAlaMetLysGluGln..................... 33
               |||   |||   :::|||
    3691 .......GTGAAGATGGTGCAGCAGTGACAGGATCTCAAACCATTAACGG 3733

      34 ..AsnGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsn 48
           |||      |||:::|||||||||:::::::::::::::::|||
    3734 TCAACACCTGTACTTTAGAGCAAACGGTGTTCAGGTCAAGGGTGAAT 3780

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-210-361-1

seq_documentation_block:
; Sequence 1, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
;  APPLICANT: Nichols, Scott E.
;  TITLE OF INVENTION: Substitutes for Modified Starches and
;  TITLE OF INVENTION: Latexes in Paper Manufacture
;  FILE REFERENCE: 0357CR
;  CURRENT APPLICATION NUMBER: US/09/210,361
;  CURRENT FILING DATE: 1998-12-11
;  EARLIER APPLICATION NUMBER: 09/007,999
;  EARLIER FILING DATE: 1998-01-16
;  EARLIER APPLICATION NUMBER: 08/478,704
;  EARLIER FILING DATE: 1995-06-07
;  EARLIER APPLICATION NUMBER: 09/009,620
;  EARLIER FILING DATE: 1998-01-20
;  EARLIER APPLICATION NUMBER: 08/485,243
;  EARLIER FILING DATE: 1995-06-07
;  EARLIER APPLICATION NUMBER: 09/008,172
;  EARLIER FILING DATE: 1998-01-16
;  EARLIER APPLICATION NUMBER: 08/482,711
;  EARLIER FILING DATE: 1995-06-07
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 4460
;   TYPE: DNA
;   ORGANISM: streptococcus mutans
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (9)...(4434)
US-09-210-361-1


alignment_scores:
            Quality:   59.00              Length:     49
              Ratio:  2.034                Gaps:      2
 Percent Similarity:  59.184   Percent Identity: 32.653

alignment_block:
 US-10-048-197-2 x US-09-210-361-1   ..
```

```
Align seg 1/1  to: US-09-210-361-1  from: 1  to: 4460

      9 ThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerTh 25
        |||:::   :::||||||   :::      |||:::||||||:::|||
   3643 ACCGTAATCGTTTTATTGAAAACGAAGAAGGTAAATGGCTGTATCTCG.. 3690

     25 rGlyAsnValAlaMetLysGluGln..................... 33
           |||   |||   :::|||
   3691 .......GTGAAGATGGTGCAGCAGTGACAGGATCTCAAACCATTAACGG 3733

     34 ..AsnGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsn 48
          |||      |||:::||||||||||:::::::::::::::::|||
   3734 TCAACACCTGTACTTTAGAGCAAACGGTGTTCAGGTCAAGGGTGAAT 3780
```

OM of: US-10-048-197-2 to: N_Geneseq_032802:* out_format : pfs

Date: Sep 18, 2002 5:51 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10048197/runat_17092002_135850_22677/app_query.fasta_1.168
-DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=std.rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10048197_@CGN1_1_523 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-10-048-197-2
Query length: 111
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 278.560000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF59778	+	558.00	1386.78	6.8e-69	3	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF28551	+	558.00	1323.03	2.4e-65	9	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV03597	+	80.50	188.55	0.0373	26	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA15394	+	80.50	188.55	0.0373	26	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA76029	+	80.00	184.60	0.0619	33	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA15300	+	80.00	184.60	0.0619	33	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA05521	+	78.00	168.53	0.4861	90	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV52208	+	78.00	143.66	11.82	81	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS62620	+	74.50	152.49	3.81	17	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF28550	-	71.00	97.74	4.3e+03	9	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL10149	-	70.00	147.58	7.14	9	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL10148	-	70.00	134.93	36.19	2	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL12634	-	70.00	133.67	42.54	3	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX13387	+	70.00	130.90	60.64	42	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF28552	+	70.00	95.09	6.0e+03	1	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF55109	+	69.50	153.51	3.34	5	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF55108	+	69.50	153.44	3.37	5	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAC85071	+	69.50	147.57	7.15	8	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC43179	+	69.50	145.81	8.96	10	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:AAT42063	+	69.50	61.08	4.4e+05	18	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAD12599	-	69.00	143.08	12.72	1	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA81456	-	69.00	101.30	2.7e+03	46	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAF21612	-	69.00	78.52	5.0e+04	34	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA81489	-	69.00	68.66	1.7e+05	83	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABA03041	-	68.50	53.19	1.1e+06	29	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC37886	-	68.00	150.73	4.77	47	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC99183	+	68.00	148.27	6.54	58	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC35309	+	68.00	140.36	18.02	11	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAD21310	+	68.00	133.58	43.04	2	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH17234	-	68.00	132.77	47.74	2	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV52205	+	68.00	111.69	712.69	14	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH15139	+	67.50	140.33	18.09	1	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ51314	+	67.50	135.95	31.75	15	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:AAT27729	+	67.50	132.55	49.07	21	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAF18124	-	67.50	126.87	101.78	34	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV19980	-	67.50	124.78	132.94	41	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ11169	-	67.50	124.78	132.94	41	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA71214	-	67.50	124.78	132.94	41	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF31912	-	67.50	124.78	132.94	4	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL26466	-	67.50	111.75	707.51	1	
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABL33364	-	67.50	108.34	1.1e+03		
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC76607	+	67.00	134.79	36.85		
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH14505	-	67.00	129.96	68.41		
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS31027	+	67.00	129.53	72.35		
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL11270	+	67.00	128.84	79.06		

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF59778

seq_documentation_block:
ID AAF59778 standard; DNA; 336 BP.
XX
AC AAF59778;
XX
DT 02-MAY-2001 (first entry)
XX
DE Moraxella catarrhalis strain ATCC43617 BASB122 DNA.
XX
KW BASB122 protein; strain ATCC43617; antigen; antibody; vaccine;
KW genetic immunisation; infection; upper respiratory tract; otitis media;
KW hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
KW invasive disease; antibacterial; auditory; ds.
XX
OS Moraxella catarrhalis.
XX
PN WO200109337-A2.
XX
PD 08-FEB-2001.
XX
PF 31-JUL-2000; 2000WO-EP07365.
XX
PR 30-JUL-1999; 99GB-0018034.
PR 30-JUL-1999; 99GB-0018036.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2001-159874/16.
DR P-PSDB; AAB60640.
XX
PT New BASB122 and BASB124 polypeptides and polynucleotides from Moraxella
PT catarrhalis strain ATCC 43617, useful as therapeutic agents or vaccines
PT against bacterial infections, e.g. otitis media or pneumonia -
XX
PS Claim 13; Page 66; 75pp; English.
XX
CC The invention relates to the Moraxella catarrhalis strain ATCC43617
CC BASB122 and BASB124 proteins (AAB60640 and AAB60641, respectively)
CC and to DNA encoding them (AAF59778 and AAF59779, respectively). The
CC invention also relates to immunogenic fragments of the BASB122 and
CC BASB124 proteins, expression vectors and host cells comprising BASB122
CC or BASB124 nucleic acids, the recombinant production of BASB122 or
CC BASB124, vaccine compositions comprising the BASB122 or BASB124 proteins
CC or nucleic acids, an antibody against BASB122 or BASB124, therapeutic
CC compositions comprising an anti-BASB122 or anti-BASB124 antibody, and a
CC method of identifying a Moraxella catarrhalis infection via the
CC detection of BASB122 or BASB124 proteins or antibodies. The vaccine
CC compositions of the invention are useful as prophylactic or therapeutic
CC agents against Moraxella catarrhalis infections in mammals, particularly
CC humans. Moraxella catarrhalis is a Gram negative bacterium frequently
CC isolated from the human upper respiratory tract, which is responsible for
CC several pathological conditions. It is responsible for about 15% of
CC otitis media cases in children (which can lead to temporary or permanent
CC hearing loss). It also causes pneumonia in elderly people, and sinusitis,
CC nosocomial infections and, less frequently, invasive diseases. BASB122 or
CC BASB124 proteins or nucleotides may additionally be used in screening for
CC novel antibacterial compounds, and in the diagnosis and staging of
CC infections. The present sequence represents DNA encoding the Moraxella
CC catarrhalis strain ATCC43617 BASB122 protein.
XX
SQ Sequence 336 BP; 106 A; 65 C; 61 G; 104 T; 0 other;

```
alignment_scores:
            Quality:  558.00              Length:    111
              Ratio:  5.027                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAF59778   ..

 Align seg 1/1  to: AAF59778  from: 1  to: 336

       1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGTTAAATCATCATATTCGCCTGACTATTTCCGCTTTACTGACGGCACT 50

      17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 TTTGGTAACAGGTTGTGTTTCTACTGGTAATGTTGCAATGAAAGAGCAAA 100

      34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 ACCAACAAACCATTGAGCAGACCATCATTAAGGGCAAGACCAATAAGCAA 150

      51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVa 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GAGATTTCTAGTAGATTTGGTTCTGCTGATAGCATCTCTTTTATGATAGT 200

      67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 GGTAATAAAATTTGGACATACCGCCATACTCGCTCCAAACCGATGGCAAG 250

      84 luIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 AAATTTTATCCCTTATAATTTCTTTTCTTTGGGTGAAACCATACAGACCA 300

     101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
         |||||||||||||||||||||||||||||||||
     301 AAGAACTTGTCATTTTATTTGACAGCAAAGGCA 333

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF28551

seq_documentation_block:
ID   AAF28551 standard; DNA; 94750 BP.
XX
AC   AAF28551;
XX
DT   04-APR-2001  (first entry)
XX
DE   Genomic fragment #38.
XX
KW   Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW   bronchopulmonary; endocarditis; meningitis; ss.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200078968-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US16649.
XX
PR   18-JUN-1999;   99US-0140121.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lagace RE,  Patterson C,  Berg KL;
XX
DR   WPI; 2001-041427/05.
XX
PT   Genomic library for identifying diagnostic and therapeutic
PT   compositions, and for identifying virulence factors, regulatory
PT   elements and drug targets, comprises Moraxella catarrhalis nucleic
PT   acids -
XX
PS   Claim 1; Page 415-436; 545pp; English.
XX
CC   The present invention relates to a Moraxella catarrhalis genomic library
CC   comprising of a combination of 41 nucleic acid molecules (see
CC   AAF28514-AAF28554). The library has a number of uses described in the
CC   specification e.g. is useful for aidentifying diagnostic and therapeutic
CC   compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC   aerobic, gram-negative diplococcus, normally found among the bacterial
CC   flora of human upper airways. M. catarrhalis is known to cause acute,
CC   localised infections such as otitis media, sinusitis and bronchopulmonary
CC   infection and life-threatening, systemic diseases including endocarditis
CC   and meningitis.
XX
SQ   Sequence 94750 BP; 27128 A; 18811 C; 21444 G; 27367 T; 0 other;


alignment_scores:
            Quality:  558.00              Length:    111
              Ratio:  5.027                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAF28551   ..

 Align seg 1/1  to: AAF28551  from: 1  to: 94750

       1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75267 ATGTTAAATCATCATATTCGCCTGACTATTTCCGCTTTACTGACGGCACT 75316

      17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75317 TTTGGTAACAGGTTGTGTTTCTACTGGTAATGTTGCAATGAAAGAGCAAA 75366

      34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75367 ACCAACAAACCATTGAGCAGACCATCATTAAGGGCAAGACCAATAAGCAA 75416

      51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVa 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75417 GAGATTTCTAGTAGATTTGGTTCTGCTGATAGCATCTCTTTTATGATAGT 75466

      67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75467 GGTAATAAAATTTGGACATACCGCCATACTCGCTCCAAACCGATGGCAAG 75516

      84 luIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75517 AAATTTTATCCCTTATAATTTCTTTTCTTTGGGTGAAACCATACAGACCA 75566

     101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
         |||||||||||||||||||||||||||||||||
   75567 AAGAACTTGTCATTTTATTTGACAGCAAAGGCA 75599

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV03597

seq_documentation_block:
ID   AAV03597 standard; DNA; 267 BP.
XX
AC   AAV03597;
XX
DT   22-OCT-1998  (first entry)
XX
DE   Neisseria meningitidis DNA sequence E29.
XX
KW   N. gonorrheae; N. lactamica; chromosome Z2491; region 1; region 2;
KW   region 3; pathogenicity; blood-brain barrier; diagnosis; infection;
KW   meningitis; ss.
XX
OS   Neisseria meningitidis.
XX
PN   WO9802547-A2.
XX
```

```
PD   22-JAN-1998.
XX
PF   11-JUL-1997;   97WO-FR01295.
XX
PR   12-JUL-1996;   96FR-0008768.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA   (SMIK ) SMITHKLINE BEECHAM.
XX
PI   Nassif X,  Tinsley C,  Achtman M,  Merker P,  Ruelle J;
PI   Vinals C;
XX
DR   WPI; 1998-110594/10.
XX
PT   Genes present in Neisseria meningitidis but not other Neisseria
PT   species - and related host cells, RNA, anti-sense sequences,
PT   polypeptide(s) and antibodies, useful for diagnosing Neisseria
PT   meningitidis infection and in protective vaccines
XX
PS   Example 4; Page 125; 150pp; French.
XX
CC   AAV03575-606 represent sequences that are present in Neisseria
CC   meningitidis and N. gonorrheae but not in N. lactamica, except for the
CC   genes involved in biosynthesis of the capsule polysaccharide, frpA or C,
CC   opc, porA, rotamase, sequence IC1106, IgA protease, pillin, pilC,
CC   proteins which bind transferrin and opacity proteins. The DNA sequences
CC   are responsible for the differences in pathogenicity between N.
CC   meningitidis and N. gonorrheae, specifically they include the genes that
CC   allow N. meningitidis to cross the blood-brain barrier. DNA sequences
CC   common to N. meningitidis and N. gonorrheae, but absent from N.
CC   lactamica, are responsible for colonisation and penetration of the
CC   mucosa. The DNA sequences can be used to produce probes and primers, and
CC   antibodies produced against the encoded proteins are used in standard
CC   hybridisation/immunoassay processes for diagnosis of N. meningitidis
CC   infection, particularly meningitis.
XX
SQ   Sequence 267 BP; 71 A; 70 C; 61 G; 65 T; 0 other;


alignment_scores:
            Quality:   80.50              Length:     71
              Ratio:  1.677                Gaps:      1
 Percent Similarity:  67.606   Percent Identity: 33.803

alignment_block:
 US-10-048-197-2 x AAV03597   ..

 Align seg 1/1  to: AAV03597  from: 1  to: 267

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||:::   :::||||||   :::::::||||||   :::::::||||||
      14 CGCATCATCGTTTCGGCTGCGTTCGTTGCGTTGGCATTAGCAGGTTGCGG 63

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
          |||   :::|||||||:::::::::::::|||:::   |||      :::
      64 CTCAATCAATAATGTAACCGTTTCCGACCAGAAACTTCAGGAA...CGTG 110

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
           ::::::      |||   ||||||      :::|||||||:::|||
     111 CCGCGTTTGCCTTGGGCGTCACCAATGCCGTAAAAATCAGCAACCGCAGC 160

      57 GlySerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHi 73
         :::::::::      |||:::|||      :::|||   |||
     161 AATGAAGGCATACGCATCAACTTTACCGCAACTGTGGGTAAGCGCGTGAC 210

      73 sThrAlaIleLeu 77
         :::|||:::|||
     211 CAATGCTATGTTA 223

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA15394

seq_documentation_block:

ID   AAA15394 standard; DNA; 267 BP.
XX
AC   AAA15394;
XX
DT   04-SEP-2000  (first entry)
XX
DE   Genomic fragment of Neisseria meningitidis Z2491.
XX
KW   Pathogenic strain; Neisseria; vaccine; Neisseria infection; ss.
XX
OS   Neisseria meningitidis.
XX
PN   WO200026375-A2.
XX
PD   11-MAY-2000.
XX
PF   28-OCT-1999;   99WO-FR02643.
XX
PR   30-OCT-1998;   98FR-0013693.
XX
PA   (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Aujame L,  Bouchardon A,  Renauld-Mongenie G,  Rokbi B,  Nassif X;
PI   Tinsley C,  Perrin A;
XX
DR   WPI; 2000-365622/31.
XX
PT   New polypeptide specific for pathogenic Neisseria useful in therapeutic
PT   or preventative vaccines and for diagnosis  -
XX
PS   Claim 1; Page 43; 187pp; French.
XX
CC   The present sequence represents a genomic fragment of Neisseria
CC   meningitidis. The specificatino describes proteins that are specific
CC   for pathogenic strains of Neisseria. The polynucleotides, polypeptides,
CC   or their antigenic fragments, are used in vaccines to treat or protect
CC   against Neisseria infections, particularly by N. meningitidis. The
CC   polynucleotide sequences are also used for recombinant production of
CC   the polypeptide and to produce attenuated Neisseria strains that
CC   overexpress it, or express it in a non-toxic mutant form.
CC   note: the present sequence is included in Claim 1, but it is those
CC   sequences that do not include the present sequence that are actually
CC   claimed.
XX
SQ   Sequence 267 BP; 71 A; 70 C; 61 G; 65 T; 0 other;


alignment_scores:
            Quality:   80.50              Length:     71
              Ratio:  1.677                Gaps:      1
 Percent Similarity:  67.606   Percent Identity: 33.803

alignment_block:
 US-10-048-197-2 x AAA15394   ..

 Align seg 1/1  to: AAA15394  from: 1  to: 267

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||:::   :::||||||   :::::::||||||   :::::::||||||
      14 CGCATCATCGTTTCGGCTGCGTTCGTTGCGTTGGCATTAGCAGGTTGCGG 63

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
          |||   :::|||||||:::::::::::::|||:::   |||      :::
      64 CTCAATCAATAATGTAACCGTTTCCGACCAGAAACTTCAGGAA...CGTG 110

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
           ::::::      |||   ||||||      :::|||||||:::|||
     111 CCGCGTTTGCCTTGGGCGTCACCAATGCCGTAAAAATCAGCAACCGCAGC 160

      57 GlySerAlaAspSerIleSerPheMetIleValValIleLysPheGlyHi 73
         :::::::::      |||:::|||      :::|||   |||
     161 AATGAAGGCATACGCATCAACTTTACCGCAACTGTGGGTAAGCGCGTGAC 210
```

```
      73 sThrAlaIleLeu 77
         :::|||:::|||
     211 CAATGCTATGTTA 223

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA76029

seq_documentation_block:
ID   AAA76029 standard; DNA; 339 BP.
XX
AC   AAA76029;
XX
DT   19-JAN-2001  (first entry)
XX
DE   Neisseria meningitidis BASB059 gene.
XX
KW   BASB059; respiratory tract infection; invasive bacterial disease;
KW   bacteraemia; meningitis; ds.
XX
OS   Neisseria meningitidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..339
FT                   /*tag= a
FT                   /product= "BASB059"
XX
PN   WO200044904-A1.
XX
PD   03-AUG-2000.
XX
PF   25-JAN-2000; 2000WO-EP00561.
XX
PR   29-JAN-1999;   99GB-0002070.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2000-505978/45.
DR   P-PSDB; AAB21107.
XX
PT   New isolated polypeptide from Neisseria meningitidis is useful for
PT   detection and treatment of N. meningitidis infection -
XX
PS   Claim 7; Page 61; 77pp; English.
XX
CC   The present sequence is the Neisseria meningitidis BASB059 gene. The
CC   gene, its protein, agonists and antagonists can be used in diagnostic
CC   assays for the bacterium, as vaccines to prevent infection and as
CC   treatments for bacterial infection, particularly those caused by
CC   Neisseria meningitidis, such as upper respiratory tract infections,
CC   invasive bacterial diseases, bacteraemia and meningitis.
XX
SQ   Sequence 339 BP; 93 A; 84 C; 84 G; 78 T; 0 other;


alignment_scores:
            Quality:   80.00            Length:     89
              Ratio:  1.455              Gaps:      1
 Percent Similarity:  61.798   Percent Identity: 29.213

alignment_block:
 US-10-048-197-2 x AAA76029   ..

 Align seg 1/1  to: AAA76029  from: 1  to: 339

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||:::   :::||||||   ::::::||||||   ::::::||||||
      10 CGCATCATCGTTTCGGCTGCGTTCGTTGCGTTGGCATTAGCAGGTTGCGG 59

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
          |||   :::||||||:::::::::::::|||:::   |||
      60 CTCAATCAATAATGTAACCGTTTCCGACCAGAAACTTCAGGAACGTGCCG 109

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
                 :::      :::   |||      :::||||||:::|||
     110 CGTTTGCCTTGGGCGTCAGCCAAAATGCCGTAAAAATCAGCAACCGCAGC 159

      57 GlySerAlaAspSerIleSerPheMetIleValVal.IleLysPheGlyH 73
         :::::::::      |||:::|||      :::||| ::::::
     160 AATGAAAGCATACGCATCAACTTTACCGCAACTGTGGGTAAGCGCGTGAG 209

      73 isThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIle 89
         :::|||:::|||               |||:::   |||   |||
     210 CCAATGCTATGTTACCAGTGTAATCAGCACAATCGGCGTTACCACTTCCG 259

      90 IleSerPheLeuTrp 94
         ::::::|||:::|||
     260 ATGCAATTTGTTTGG 274

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA15300

seq_documentation_block:
ID   AAA15300 standard; DNA; 339 BP.
XX
AC   AAA15300;
XX
DT   04-SEP-2000  (first entry)
XX
DE   DNA encoding a polypeptide of a Neisseria pathogenic strain.
XX
KW   Pathogenic strain; Neisseria; vaccine; Neisseria infection; ss.
XX
OS   Neisseria meningitidis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..339
FT                   /*tag= a
XX
PN   WO200026375-A2.
XX
PD   11-MAY-2000.
XX
PF   28-OCT-1999;   99WO-FR02643.
XX
PR   30-OCT-1998;   98FR-0013693.
XX
PA   (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Aujame L,  Bouchardon A,  Renauld-Mongenie G,  Rokbi B,  Nassif X;
PI   Tinsley C,  Perrin A;
XX
DR   WPI; 2000-365622/31.
DR   P-PSDB; AAY93269.
XX
PT   New polypeptide specific for pathogenic Neisseria useful in therapeutic
PT   or preventative vaccines and for diagnosis  -
XX
PS   Claim 2; Page 69; 187pp; French.
XX
CC   The present sequence encodes a protein that is specific for pathogenic
CC   strains of Neisseria. The polynucleotides, polypeptides, or their
CC   antigenic fragments, are used in vaccines to treat or protect against
CC   Neisseria infections, particularly by N. meningitidis. The
CC   polynucleotide sequence is also used for recombinant production of
CC   the polypeptide and to produce attenuated Neisseria strains that
CC   overexpress it, or express it in a non-toxic mutant form.
XX
SQ   Sequence 339 BP; 93 A; 84 C; 84 G; 78 T; 0 other;


alignment_scores:
            Quality:   80.00            Length:     89
              Ratio:  1.455              Gaps:      1
 Percent Similarity:  61.798   Percent Identity: 29.213
```

```
alignment_block:
 US-10-048-197-2 x AAA15300   ..

 Align seg 1/1  to: AAA15300  from: 1  to: 339

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||:::   :::||||||   ::::::||||||   ::::::||||||
      10 CGCATCATCGTTTCGGCTGCGTTCGTTGCGTTGGCATTAGCAGGTTGCGG 59

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
          |||   :::||||||:::::::::::::|||:::   |||
      60 CTCAATCAATAATGTAACCGTTTCCGACCAGAAACTTCAGGAACGTGCCG 109

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
                :::      :::   |||      :::||||||:::|||
     110 CGTTTGCCTTGGGCGTCAGCCAAAATGCCGTAAAAATCAGCAACCGCAGC 159

      57 GlySerAlaAspSerIleSerPheMetIleValVal.IleLysPheGlyH 73
         :::::::::      |||:::|||      :::||| ::::::
     160 AATGAAAGCATACGCATCAACTTTACCGCAACTGTGGGTAAGCGCGTGAG 209

      73 isThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIle 89
           :::|||:::|||               |||:::   |||   |||
     210 CCAATGCTATGTTACCAGTGTAATCAGCACAATCGGCGTTACCACTTCCG 259

      90 IleSerPheLeuTrp 94
         ::::::|||:::|||
     260 ATGCAATTTGTTTGG 274

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA05521

seq_documentation_block:
ID   AAA05521 standard; DNA; 900 BP.
XX
AC   AAA05521;
XX
DT   24-MAY-2000  (first entry)
XX
DE   Streptococcus pneumoniae type 4 nucleotide sequence 4170.6.
XX
KW   Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW   antibacterial; antiinflammatory; meningitis; infection; diagnosis;
KW   pneumococcal disease; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200006737-A2.
XX
PD   10-FEB-2000.
XX
PF   27-JUL-1999;   99WO-GB02451.
XX
PR   27-JUL-1998;   98GB-0016337.
PR   19-MAR-1999;   99US-0125164.
XX
PA   (MICR-) MICROBIAL TECHNICS LTD.
XX
PI   Gilbert CFG,  Hansbro PM;
XX
DR   WPI; 2000-195300/17.
XX
PT   New Streptococcal protein, useful as a vaccine, for diagnosis of
PT   pneumococcal diseases and for screening agents capable of antagonizing
PT   or inhibiting expression of the protein
XX
PS   Claim 7; Page 54; 108pp; English.
XX
CC   AAY81501 to AAY81679 represent specifically claimed protein sequences
CC   isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
CC   specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC   The sequences have antibacterial and antiinflammatory properties.
CC   The protein sequences, and fragments of them, are useful as immunogens
CC   and/or antigens. The nucleotide sequences can be used in vaccines and in
CC   diagnostic assays. The proteins and nucleotides can be useful for the
CC   detection and diagnosis of S. pneumoniae. The protein sequences are also
CC   useful for screening an agent capable of antagonising, inhibiting or
CC   interfering with the function or expression of the proteins in which the
CC   agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC   and meningitis. AAA05591 to AAA05614 represent primers used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 900 BP; 240 A; 155 C; 211 G; 294 T; 0 other;


alignment_scores:
           Quality:   78.00            Length:    127
             Ratio:  1.200              Gaps:      6
 Percent Similarity:  51.181   Percent Identity: 28.346

alignment_block:
 US-10-048-197-2 x AAA05521   ..

 Align seg 1/1  to: AAA05521  from: 1  to: 900

      10 IleSerAlaLeuLeuThrAla...LeuLeuValThrGlyCysValSerTh 25
         ::::::||||||:::||||||   :::||||||||||||   |||   ::
     232 CTAGGAGCCTTGGTAACAGCTGTGATTCTCGTAACGGGCTCTGTTCTAGT 281

      25 r......GlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
         :         |||||||:::         :::   |||   :::   ::::
     282 CATTTTGGAAAATGTCACGAAGATTTTGCATCCGCAACCAGTCAATGATG 331

      40 lnThrIleIle..................................... 43
         ::   |||:::
     332 AGGGGATTCTCTGGTTAGGAATTATTGCGATTACTATCAATCTGTTAGCG 381

      44 ..............LysGlyLysThrAsnLysGlnGluIleSerSerAr 55
                       ||||||||||||:::::::::::::|||   |||
     382 AGTCTGGTGGTTGGTAAGGGAAAGACAAAGAATGAGTCTATTCTGAGTCT 431

      55 gPheGlySerAlaAspSerIleSerPhe..............MetIleV 67
                    |||:::::::::::::                |||   :
     432 GCATTTTCTGGAAGATACGCTAGGGTGGGTAGCTGTTATCCTGATGGCGA 481

      67 alValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 83
         ::|||:::::|||            ||||||   |||
     482 TTGTTCTTCGATTTACGGACTGGTATATCCTAGATCCT............ 519

      84 GluIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPr 100
            :::||||||||||:::||||||||||:::   ::::::         ||
     520 ...CTTTTGTCCCTTGTCATTTCTTTCTTTATTCTTTCAAAAGCCCTTCC 566

     100 oLysAsnLeuSerPheTyrLeuThrAlaLys 110
         |:::         |||:::   |||   |||
     567 ACGT.........TTTTGGTCTACACTCAAG 588

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV52208

seq_documentation_block:
ID   AAV52208 standard; DNA; 8136 BP.
XX
AC   AAV52208;
XX
DT   23-OCT-1998  (first entry)
XX
DE   Streptococcus pneumoniae genome fragment SEQ ID NO:75.
XX
KW   Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW   computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9818931-A2.
XX
PD   07-MAY-1998.
```

```
XX
PF   30-OCT-1997;   97WO-US19588.
XX
PR   31-OCT-1996;   96US-0029960.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Dougherty BA,  Fannon M;
PI   Kunsch CA,  Rosen CA;
XX
DR   WPI; 1998-272225/24.
XX
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
XX
PS   Claim 1; Page 617-622; 1409pp; English.
XX
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC   recorded on it, or a representative fragment or a sequence at least 95%
CC   identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC   SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC   Streptococcus pneumoniae. The present invention also describes an
CC   isolated nucleic acid molecule encoding a homologue of any of the
CC   fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC   nucleic acid molecule is produced by a process comprising: (a) screening
CC   a genomic DNA library using as a probe a target sequence defined by any
CC   of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC   library which contain sequences that hybridise to the target sequence and
CC   isolating the nucleic acid molecules from the members; or (b) isolating
CC   mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC   molecules whose nucleotide sequence is homologous to amplification
CC   primers derived from the fragment of the S. pneumoniae genome to prime
CC   the amplification and isolating the amplified sequences. The computer
CC   readable medium can be used in a computer-based system for identifying
CC   fragments of the S. pneumoniae genome of commercial importance, or
CC   expression modulating fragments of the S. pneumoniae genome. Products
CC   from the present invention can be used in diagnosis kits and assays, and
CC   pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ   Sequence 8136 BP; 2249 A; 1481 C; 1983 G; 2423 T; 0 other;


alignment_scores:
            Quality:   78.00              Length:    127
              Ratio:  1.200                Gaps:      6
 Percent Similarity:  51.181   Percent Identity: 28.346

alignment_block:
 US-10-048-197-2 x AAV52208   ..

 Align seg 1/1  to: AAV52208  from: 1  to: 8136

      10 IleSerAlaLeuLeuThrAla...LeuLeuValThrGlyCysValSerTh 25
         ::::::||||||:::||||||   :::||||||||||||   |||   ::
    1926 CTAGGAGCCTTGGTAACAGCTGTGATTCTCGTAACGGGCTCTGTTCTAGT 1975

      25 r......GlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
         :        ||||||:::          :::   |||   :::   ::::
    1976 CATTTTGGAAAATGTCACGAAGATTTTGCATCCGCAACCAGTCAATGATG 2025

      40 lnThrIleIle..................................... 43
         ::   |||:::
    2026 AGGGGATTCTCTGGTTAGGAATTATTGCGATTACTATCAATCTGTTAGCG 2075

      44 ..............LysGlyLysThrAsnLysGlnGluIleSerSerAr 55
                       ||||||||||||:::::::::::::|||   |||
    2076 AGTCTGGTGGTTGGTAAGGGAAAGACAAAGAATGAGTCTATTCTGAGTCT 2125

      55 gPheGlySerAlaAspSerIleSerPhe..............MetIleV 67
                     |||::::::::::::              |||   :

    2126 GCATTTTCTGGAAGATACGCTAGGGTGGGTAGCTGTTATCCTGATGGCGA 2175

      67 alValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 83
         ::|||::::::|||          ||||||   |||
    2176 TTGTTCTTCGATTTACGGACTGGTATATCCTAGATCCT............ 2213

      84 GluIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPr 100
            :::||||||||||:::|||||||||:::   ::::::         ||
    2214 ...CTTTTGTCCCTTGTCATTTCTTTCTTTATTCTTTCAAAAGCCCTTCC 2260

     100 oLysAsnLeuSerPheTyrLeuThrAlaLys 110
         |:::         |||:::   |||   |||
    2261 ACGT.........TTTTGGTCTACACTCAAG 2282

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:AAS62620

seq_documentation_block:
ID   AAS62620 standard; cDNA; 1709 BP.
XX
AC   AAS62620;
XX
DT   14-FEB-2002  (first entry)
XX
DE   cDNA sequence #407 encoding novel human secreted protein.
XX
KW   Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW   immune deficiency disorder; blood disorder; inflammatory disorder;
KW   infectious disorder; gene therapy; antimicrobial; hepatotropic;
KW   immunosuppressive; antirheumatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200177291-A2.
XX
PD   18-OCT-2001.
XX
PF   29-MAR-2001; 2001WO-US10485.
XX
PR   06-APR-2000; 2000US-195604P.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Wong GG,  Clark HF,  Fechtel K,  Agostino MJ,  Howes SH,  Resnick RJ;
PI   Gulukota K,  Graham JR;
XX
DR   WPI; 2002-010900/01.
XX
PT   New polynucleotides encoding secreted proteins useful for treating e.g.
PT   asthma, HIV and Crohn's disease -
XX
PS   Claim 1; Page 289-290; 391pp; English.
XX
CC   The present invention relates to the isolation of novel cDNA sequences
CC   which encode human secreted proteins. The cDNA sequences have been
CC   derived from a variety of human tissues. The invention also provides
CC   a method for producing proteins from these polynucleotide sequences.
CC   The proteins are useful for identifying compounds that modulate their
CC   activity and production, and the cell is also useful for identifying
CC   compounds that modulate expression of the polynucleotide sequences
CC   encoding the secreted proteins. The sequences of the invention are
CC   useful for treating diseases such as hyperproliferative disorders
CC   (e.g. cancer), immune deficiency disorders (e.g. severe combined
CC   immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC   sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory
CC   disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis).
CC   The polynucleotide sequences of the invention are also useful in gene
CC   therapy. AAS62214-AAS62838 represent the cDNA sequences of the
CC   invention that encode for novel human secreted proteins.
XX
SQ   Sequence 1709 BP; 481 A; 358 C; 421 G; 449 T; 0 other;


alignment_scores:
```

```
           Quality:   74.50            Length:    126
             Ratio:  1.112              Gaps:      4
 Percent Similarity:  53.175   Percent Identity: 21.429

alignment_block:
 US-10-048-197-2 x AAS62620   ..

 Align seg 1/1  to: AAS62620  from: 1  to: 1709

      10 IleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCys..... 22
         :::|||:::   :::::::::   |||::::::   |||
     224 CTGAGCGTGAACGTGGTATCACCATTGATATCTCCTTGTGGAAATTTGAG 273

      23 ....ValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrI 38
            :::||||||      :::::::::      |||:::   :::|||:
     274 ACCAGCAAGTACTATGTGACTATCATTGATGCCCCAGGACACAGAGACTT 323

      38 leGluGlnThrIleIleLysGlyLys..................... 46
         :::::::::|||   :::::::|||
     324 TATCAAAAACATGATTACAGGGACATCTCAGGCTGACTGTGCTGTCCTGA 373

      47 .................ThrAsnLysGlnGluIleSerSerArgPheGl 57
                         :::|||   :::   :::|||   |||:::||
     374 TTGTTGCTGCTGGTGTTGGTGAATTTGAAGCTGGTATCTCCAAGAATGGG 423

      57 ySerAlaAspSerIleSerPheMetIle..................... 66
         |      :::|||:::   |||   :::
     424 CAGACCCGAGAGCATGCCCTTCTGGCTTACACACTGGGTGTGAAACAACT 473

      67 ........ValValIleLysPheGlyHisThrAlaIleLeuAlaPro.As 80
                |||:::   |||:::   |||   |||   |||   ||| ::
     474 AATTGTCGGTGTTAACAAAATGGATTCACTGAGCCACCCTACAGCCAGAA 523

      80 nArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrpValLysP 97
         :|||::::::||||||::::::   :::   :::::::::
     524 GAGATATGAGGAAATTGTTAAGGAAGTCAGCACTTACATTAAGAAAATTG 573

      97 roTyrArgProLysAsnLeuSerPhe 105
            |||:::|||   :::::::::|||
     574 GCTACAACCCCGACACAGTAGCATTT 599

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF28550

seq_documentation_block:
ID   AAF28550 standard; DNA; 99629 BP.
XX
AC   AAF28550;
XX
DT   04-APR-2001  (first entry)
XX
DE   Genomic fragment #37.
XX
KW   Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW   bronchopulmonary; endocarditis; meningitis; ss.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200078968-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US16649.
XX
PR   18-JUN-1999;   99US-0140121.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lagace RE,  Patterson C,  Berg KL;
XX
DR   WPI; 2001-041427/05.
XX
PT   Genomic library for identifying diagnostic and therapeutic
PT   compositions, and for identifying virulence factors, regulatory
PT   elements and drug targets, comprises Moraxella catarrhalis nucleic
PT   acids -
XX
PS   Claim 1; Page 391-415; 545pp; English.
XX
CC   The present invention relates to a Moraxella catarrhalis genomic library
CC   comprising of a combination of 41 nucleic acid molecules (see
CC   AAF28514-AAF28554). The library has a number of uses described in the
CC   specification e.g. is useful for aidentifying diagnostic and therapeutic
CC   compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC   aerobic, gram-negative diplococcus, normally found among the bacterial
CC   flora of human upper airways. M. catarrhalis is known to cause acute,
CC   localised infections such as otitis media, sinusitis and bronchopulmonary
CC   infection and life-threatening, systemic diseases including endocarditis
CC   and meningitis.
XX
SQ   Sequence 99629 BP; 29233 A; 19222 C; 21909 G; 29264 T; 1 other;


alignment_scores:
           Quality:   71.00            Length:    109
             Ratio:  1.224              Gaps:      4
 Percent Similarity:  53.211   Percent Identity: 22.936

alignment_block:
 US-10-048-197-2 x AAF28550/rev   ..

 Align seg 1/1  to reverse of: AAF28550  from: 1  to: 99629

       4 HisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValTh 20
         :::::::|||      |||::::::::::::||||||   |||:::   :::
   76708 TATCAAATTTCCACGACCCTGAACTCACTCCTTCAGGCAGTGTCAATGAT 76659

      20 rGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnT 37
          :::   :::::::|||   |||::::::::::   :::||||||:::
   76658 AAGCGTTCTTGATACACATAATGCTTCGGTGTAACAACAAAATTCACTAC 76609

      37 hrIleGluGlnThrIle......IleLysGlyLysThrAsnLysGlnGlu 51
           :::::::::   :::       :::|||      |||      ::::::
   76608 AAGTTCATCACCCACTTGAGATGATGAAACCTTAAACGCTTCATGAACAG 76559

      52 IleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValVa 68
         ::::::      |||
   76558 GCAAAAACTTTGGCTGGC.............................. 76541

      68 lIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluI 85
                    :::      |||      ||||||:::||||||   :
   76540 ..........TGGTATCCCCAAATAGCGATGCCAAACCATTGGCAGACGC 76501

      85 leLeu.........SerLeuIleIleSerPheLeuTrpValLysProTyr 98
         :::::        :::||||||            |||:::::::|||
   76500 TGGTAAGGCTACGCCAACTGATACCGCACCGACAATGGATACGCCCA... 76454

      99 ArgProLysAsnLeuSerPheTyrLeu 107
         :::|||   |||   |||   :::|||
   76453 AAGCCATAGAATAAATCCAATTTTTTA 76427

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL10149

seq_documentation_block:
ID   ABL10149 standard; cDNA; 969 BP.
XX
AC   ABL10149;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 24929.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
```

```
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB66046.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 24929; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 969 BP; 271 A; 228 C; 249 G; 221 T; 0 other;


alignment_scores:
            Quality:   70.00              Length:     41
              Ratio:  2.500                Gaps:      2
 Percent Similarity:  68.293   Percent Identity: 43.902

alignment_block:
 US-10-048-197-2 x ABL10149/rev   ..

 Align seg 1/1  to reverse of: ABL10149  from: 1  to: 969

      54 SerArgPheGlySerAlaAspSerIleSerPheMetIleValValIleLy 70
         |||||||||   :::   |||:::::::::   :::::::::   |||
     882 AGTCGTTTCCCGGCCATCGACGCATTCGATGTAGTGGTGGCATCCATCGA 833

      70 sPheGlyHis...ThrAlaIleLeuAlaProAsnArgTrpGlnGluIleL 86
          |||||||||      ||||||||||||||||               |||
     832 TTTCGGCCACTTCGAAGCCATCCTTGCGCCC.............ATCA 798

      86 euSerLeuIleIleSerPheLeu 93
           |||:::::|||:::||||||
     797 CATCTATCGTAATCGCATTTCTG 775

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL10148

seq_documentation_block:
ID   ABL10148 standard; cDNA; 2969 BP.
XX
AC   ABL10148;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 24926.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
```

```
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB66045.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 24926; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2969 BP; 878 A; 613 C; 618 G; 860 T; 0 other;


alignment_scores:
            Quality:   70.00              Length:     41
              Ratio:  2.500                Gaps:      2
 Percent Similarity:  68.293   Percent Identity: 43.902

alignment_block:
 US-10-048-197-2 x ABL10148/rev   ..

 Align seg 1/1  to reverse of: ABL10148  from: 1  to: 2969

      54 SerArgPheGlySerAlaAspSerIleSerPheMetIleValValIleLy 70
         |||||||||   :::   |||:::::::::   :::::::::   |||
    1882 AGTCGTTTCCCGGCCATCGACGCATTCGATGTAGTGGTGGCATCCATCGA 1833

      70 sPheGlyHis...ThrAlaIleLeuAlaProAsnArgTrpGlnGluIleL 86
          |||||||||      ||||||||||||||||               |||
    1832 TTTCGGCCACTTCGAAGCCATCCTTGCGCCC.............ATCA 1798

      86 euSerLeuIleIleSerPheLeu 93
           |||:::::|||:::||||||
    1797 CATCTATCGTAATCGCATTTCTG 1775

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL12634

seq_documentation_block:
ID   ABL12634 standard; cDNA; 3319 BP.
XX
AC   ABL12634;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 32384.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
```

```
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB68531.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 32384; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3319 BP; 928 A; 721 C; 758 G; 912 T; 0 other;


alignment_scores:
            Quality:   70.00              Length:     41
              Ratio:  2.500                Gaps:      2
 Percent Similarity:  68.293    Percent Identity: 43.902

alignment_block:
 US-10-048-197-2 x ABL12634/rev   ..

 Align seg 1/1  to reverse of: ABL12634  from: 1  to: 3319

      54 SerArgPheGlySerAlaAspSerIleSerPheMetIleValValIleLy 70
         |||||||||   :::   |||:::::::::   :::::::::   |||
     570 AGTCGTTTCCCGGCCATCGACGCATTCGATGTAGTGGTGGCATCCATCGA 521

      70 sPheGlyHis...ThrAlaIleLeuAlaProAsnArgTrpGlnGluIleL 86
         |||||||||      ||||||||||||||               |||
     520 TTTCGGCCACTTCGAAGCCATCCTTGCGCCC..............ATCA 486

      86 euSerLeuIleIleSerPheLeu 93
         |||:::::::|||:::|||||||
     485 CATCTATCGTAATCGCATTTCTG 463

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX13387

seq_documentation_block:
ID   AAX13387 standard; DNA; 4239 BP.
XX
AC   AAX13387;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:450.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 1681-1683; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 4239 BP; 1556 A; 477 C; 901 G; 1292 T; 13 other;


alignment_scores:
            Quality:   70.00              Length:     96
              Ratio:  1.148                Gaps:      3
 Percent Similarity:  63.542    Percent Identity: 26.042

alignment_block:
 US-10-048-197-2 x AAX13387   ..

 Align seg 1/1  to: AAX13387  from: 1  to: 4239

       6 IleArgLeuThrIleSerAla......LeuLeuThrAlaLeuLeuValTh 20
         :::   ::::::|||:::|||      |||:::|||   |||:::   ::
     913 GTAGGTATTTCTATAGGTGCAATTGTGCTAATTACATTGCTTATTAAAGC 962

      20 rGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnT 37
         ::::   ::::::::::|||     :::      ||||||:::|||
     963 TTCAATTATTAATTCGGGGTATTTTTCTAAAGTGGAACAACGTCAAGTAT 1012

      37 hrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSer 53
           ::::::   |||   :::|||      :::|||:::::::::::::
    1013 TAGCACATATGATTATTGACAACGGGTATTACACAAAAAAACAAGTGAAA 1062

      54 SerArgPheGly...SerAlaAspSerIleSerPheMetIleValValIl 69
         |||      |||   ::::::::::::|||:::|||      :::
    1063 AGTWSTGAYGGYAAAACAARAGAAAAAATCAAGTTTCCAAAAATTTATTA 1112
```

```
      69 eLysPheGlyHisThrAlaIleLeu..............AlaProAsnA 81
         |||   :::   ::::::|||:::               |||   |||:
    1113 TAAGTCTGCAAAAAATAGTATTTTYGTTTCATTTGAAACAGCRGGTAATA 1162

      81 rgTrpGlnGluIleLeuSerLeuIleIleSerPheLeu 93
         :::::||||||   ::::::   |||   :::||||||
    1163 AATTTCAGGAAAAATTTGAAACCATTGGTGGCTTTTTG 1200

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF28552

seq_documentation_block:
ID   AAF28552 standard; DNA; 100848 BP.
XX
AC   AAF28552;
XX
DT   04-APR-2001  (first entry)
XX
DE   Genomic fragment #39.
XX
KW   Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW   bronchopulmonary; endocarditis; meningitis; ss.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200078968-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US16649.
XX
PR   18-JUN-1999;   99US-0140121.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lagace RE,  Patterson C,  Berg KL;
XX
DR   WPI; 2001-041427/05.
XX
PT   Genomic library for identifying diagnostic and therapeutic
PT   compositions, and for identifying virulence factors, regulatory
PT   elements and drug targets, comprises Moraxella catarrhalis nucleic
PT   acids -
XX
PS   Claim 1; Page 436-459; 545pp; English.
XX
CC   The present invention relates to a Moraxella catarrhalis genomic library
CC   comprising of a combination of 41 nucleic acid molecules (see
CC   AAF28514-AAF28554). The library has a number of uses described in the
CC   specification e.g. is useful for aidentifying diagnostic and therapeutic
CC   compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC   aerobic, gram-negative diplococcus, normally found among the bacterial
CC   flora of human upper airways. M. catarrhalis is known to cause acute,
CC   localised infections such as otitis media, sinusitis and bronchopulmonary
CC   infection and life-threatening, systemic diseases including endocarditis
CC   and meningitis.
XX
SQ   Sequence 100848 BP; 28518 A; 19877 C; 22976 G; 29477 T; 0 other;


alignment_scores:
            Quality:   70.00            Length:    153
              Ratio:  0.986              Gaps:     10
 Percent Similarity:  46.405   Percent Identity: 27.451

alignment_block:
 US-10-048-197-2 x AAF28552   ..

 Align seg 1/1  to: AAF28552  from: 1  to: 100848

      13 LeuLeuThrAlaLeuLeuValThrGlyCysValSerThr..... 25
         ::::::      :::|||:::::::||||||:::::|||
   50328 ATAATGATGAGATTTTTATTGGTTGGATGCTTGGCAACCTTATTTTTACT 50377

      26 .......GlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
                |||||||   :::   :::|||   :::   :::||||||||||
   50378 TGGGTGTGGTAATAATTCTAGTCAAGAATTAGAGACGAATACCATTGAAC 50427

      40 lnThr............IleIleLys..................... 44
         ||:::                |||||
   50428 AATCAAGCCCAAGAATTACAATAAAAAATGAAATGATCGAAGCTCAACCG 50477

      45 ............GlyLysThrAsnLysGln................... 50
                   |||   |||||||||
   50478 TTGCCAGAGACAGGGGCAACCAACAAACCTAATTTGTATGGCGATTCACC 50527

      51 ....GluIleSerSerArgPheGlySer..................... 58
             :::|||::::::::::   ||||||
   50528 TTTACAAATAAAAACCAACTCAGGAAGTCATTATTGGGTAAAAATTGTCA 50577

      59 ................AlaAspSerIleSerPheMetIle.ValValIl 69
                              :::   :::|||::::::||| ||||||
   50578 ATGCTTATGACGAGCGAGAAGAGCTGGTGAGTTATTTTATCCGTGGTGGA 50627

      69 eLysPheGlyHisThrAlaIleLeuAlaPro..................A 80
         |||      |||      :::|||:::|||                   |
   50628 GAAA......CATTAGATGTTACTCTGCCCGTGGGAAGTTATGTCGTGAA 50671

      80 snArgTrpGlnGluIleLeu............SerLeuIleIleSerPhe 92
         ||      :::|||||||||           |||:::   ::::::
   50672 ATATGCTTATGGAGATACTTGGTATGGCGAAGAGTATTTATTTGGCGAAA 50721

      93 LeuTrpValLysProTyrArgProLysAsnLeuSerPheTyrLeuThrAl 109
         :::   |||   |||   |||   |||   |||:::|||   :::|||
   50722 ATACAGGTTATTCCAAGGCGGATGAAGTGTTTGAATTTT...ATCACAAT 50768

     109 aLysAla 111
         ||||||
   50769 CAAGGCT 50775

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF55109

seq_documentation_block:
ID   AAF55109 standard; DNA; 513 BP.
XX
AC   AAF55109;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Nucleotide sequence of a BASB119 polypeptide.
XX
KW   BASB119; immune response; bacterial infection; genetic immunization;
KW   otitis media; pneumonia; sinusitis; nosocomial infection;
KW   invasive disease; hearing loss; fluid accumulation; antibacterial; ss.
XX
OS   Moraxella catarrhalis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..513
FT                   /*tag=  a
FT                   /product= "BASB119"
FT                   /note= "no termination codon given"
XX
PN   WO200109336-A1.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07363.
XX
PR   03-AUG-1999;   99GB-0018302.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2001-159873/16.
DR   P-PSDB; AAB67488.
```

```
XX
PT   New BASB119 polypeptides and polynucleotides from Moraxella catarrhalis
PT   strain ATCC 43617, useful as therapeutic agents or vaccines against
PT   bacterial infections, e.g. otitis media or pneumonia -
XX
PS   Claim 13; Page 65; 83pp; English.
XX
CC   The present sequence encodes a BASB119 polypeptide of Moraxella
CC   catarrhalis strain ATCC43617. BASB119 polypeptides and polynucleotides
CC   are useful for generating an immune response in an animal. The
CC   polypeptides may also be used as prophylactic agents of bacterial
CC   infections, particularly M. catarrhalis infections in mammals,
CC   especially humans. The polynucleotides are useful in therapy or
CC   prophylaxis, particularly genetic immunization against these infections
CC   or diseases. These diseases include otitis media in infants or
CC   children, pneumonia in elderlies, sinusitis, nosocomial infections and
CC   invasive diseases, chronic otitis media with hearing loss, fluid
CC   accumulation in the middle ear, infection of the upper respiratory
CC   tract, or inflammation of the middle ear. The polypeptides or
CC   polynucleotides may also be employed as research reagents and materials
CC   for discovering treatments of and diagnostics for diseases,
CC   particularly human diseases. In particular, the polypeptides or
CC   polynucleotides are useful in the discovery and development of
CC   antibacterial compounds, or for diagnosing diseases, staging of the
CC   disease, determining the response of an infectious organism to drugs.
XX
SQ   Sequence 513 BP; 173 A; 78 C; 111 G; 151 T; 0 other;


alignment_scores:
            Quality:   69.50              Length:    154
              Ratio:  0.952                Gaps:     10
 Percent Similarity:  47.403   Percent Identity: 27.922

alignment_block:
 US-10-048-197-2 x AAF55109   ..

 Align seg 1/1  to: AAF55109  from: 1  to: 513

       6 IleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCy 22
         :::|||:::   :::   :::   |||::::::|||::::::   |||||
       4 ATGAGATTTTTATTGGTTGGATGCTTGGCAACCTTATTTTTACTTGGGTG 53

      22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleG 39
         |         ||||||   :::   :::|||   :::   :::|||||||
      54 T.........GGTAATAATTCTAGTCAAGAATTAGAGACGAATACCATTG 94

      39 luGlnThr............IleIleLys..................... 44
         ||||||:::                ||||||
      95 AACAATCAAGCCCAAGAATTACAATAAAAAATGAAATGATCGAAGCTCAA 144

      45 ..............GlyLysThrAsnLysGln................... 50
                       |||   |||||||||
     145 CCGTTGCCAGAGACAGGGGCAACCAACAAACCTAATTTGTATGGCGATTC 194

      51 .......GluIleSerSerArgPheGlySer..................... 58
               :::|||:::::::::   ||||||
     195 ACCTTTACAAATAAAAACCAACTCAGGAAGTCATTATTGGGTAAAAATTG 244

      59 ...................AlaAspSerIleSerPheMetIle.ValVa 68
                             :::   :::|||::::::||| |||||
     245 TCAATGCTTATGACGAGCGAGAAGAGCTGGTGAGTTATTTTATCCGTGGT 294

      68 lIleLysPheGlyHisThrAlaIleLeuAlaPro................... 79
         |   |||      |||      :::|||:::|||
     295 GGAGAAA......CATTAGATGTTACTCTGCCCGTGGGAAGTTATGTCGT 338

      80 ..AsnArgTrpGlnGluIleLeu............SerLeuIleIleSer 91
           |||      :::|||||||||            |||:::   ::::::
     339 GAAATATGCTTATGGAGATACTTGGTATGGCGAAGAGTATTTATTTGGCG 388

      92 PheLeuTrpValLysProTyrArgProLysAsnLeuSerPheTyrLeuTh 108
            :::   |||   |||   |||   |||   |||:::|||   :::||
```

```
     389 AAAATACAGGTTATTCCAAGGCGGATGAAGTGTTTGAATTTT...ATCAC 435

     108 rAlaLysAla 111
            |   ||||||
     436 AATCAAGGCT 445

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF55108

seq_documentation_block:
ID   AAF55108 standard; DNA; 516 BP.
XX
AC   AAF55108;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Nucleotide sequence of a BASB119 polypeptide.
XX
KW   BASB119; immune response; bacterial infection; genetic immunization;
KW   otitis media; pneumonia; sinusitis; nosocomial infection;
KW   invasive disease; hearing loss; fluid accumulation; antibacterial; ss.
XX
OS   Moraxella catarrhalis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..516
FT                   /*tag= a
FT                   /product= "BASB119"
XX
PN   WO200109336-A1.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07363.
XX
PR   03-AUG-1999;   99GB-0018302.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2001-159873/16.
DR   P-PSDB; AAB67487.
XX
PT   New BASB119 polypeptides and polynucleotides from Moraxella catarrhalis
PT   strain ATCC 43617, useful as therapeutic agents or vaccines against
PT   bacterial infections, e.g. otitis media or pneumonia -
XX
PS   Claim 13; Page 65; 83pp; English.
XX
CC   The present sequence encodes a BASB119 polypeptide of Moraxella
CC   catarrhalis strain ATCC43617. BASB119 polypeptides and polynucleotides
CC   are useful for generating an immune response in an animal. The
CC   polypeptides may also be used as prophylactic agents of bacterial
CC   infections, particularly M. catarrhalis infections in mammals,
CC   especially humans. The polynucleotides are useful in therapy or
CC   prophylaxis, particularly genetic immunization against these infections
CC   or diseases. These diseases include otitis media in infants or
CC   children, pneumonia in elderlies, sinusitis, nosocomial infections and
CC   invasive diseases, chronic otitis media with hearing loss, fluid
CC   accumulation in the middle ear, infection of the upper respiratory
CC   tract, or inflammation of the middle ear. The polypeptides or
CC   polynucleotides may also be employed as research reagents and materials
CC   for discovering treatments of and diagnostics for diseases,
CC   particularly human diseases. In particular, the polypeptides or
CC   polynucleotides are useful in the discovery and development of
CC   antibacterial compounds, or for diagnosing diseases, staging of the
CC   disease, determining the response of an infectious organism to drugs.
XX
SQ   Sequence 516 BP; 174 A; 78 C; 112 G; 152 T; 0 other;


alignment_scores:
            Quality:   69.50              Length:    154
```

```
            Ratio:  0.952              Gaps:     10
 Percent Similarity:  47.403   Percent Identity: 27.922

alignment_block:
 US-10-048-197-2 x AAF55108   ..

 Align seg 1/1  to: AAF55108  from: 1  to: 516

       6 IleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCy 22
         :::|||:::   :::   :::   |||::::::|||::::::   |||||
       4 ATGAGATTTTTATTGGTTGGATGCTTGGCAACCTTATTTTTACTTGGGTG 53

      22 sValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleG 39
         |         ||||||   :::   :::|||   :::   :::|||||||
      54 T.........GGTAATAATTCTAGTCAAGAATTAGAGACGAATACCATTG 94

      39 luGlnThr............IleIleLys..................... 44
         |||||:::             ||||||
      95 AACAATCAAGCCCAAGAATTACAATAAAAAATGAAATGATCGAAGCTCAA 144

      45 ..............GlyLysThrAsnLysGln................. 50
                       |||   |||||||||
     145 CCGTTGCCAGAGACAGGGGCAACCAACAAACCTAATTTGTATGGCGATTC 194

      51 .......GluIleSerSerArgPheGlySer.................. 58
              :::|||:::::::::   ||||||
     195 ACCTTTACAAATAAAAACCAACTCAGGAAGTCATTATTGGGTAAAAATTG 244

      59 ...................AlaAspSerIleSerPheMetIle.ValVa 68
                            :::   :::|||::::::|||| |||||
     245 TCAATGCTTATGACGAGCGAGAAGAGCTGGTGAGTTATTTTATCCGTGGT 294

      68 lIleLysPheGlyHisThrAlaIleLeuAlaPro................ 79
         |   |||      |||      :::|||:::|||
     295 GGAGAAA......CATTAGATGTTACTCTGCCCGTGGGAAGTTATGTCGT 338

      80 ..AsnArgTrpGlnGluIleLeu............SerLeuIleIleSer 91
           |||      :::|||||||||           |||:::   ::::::
     339 GAAATATGCTTATGGAGATACTTGGTATGGCGAAGAGTATTTATTTGGCG 388

      92 PheLeuTrpValLysProTyrArgProLysAsnLeuSerPheTyrLeuTh 108
           :::   |||   |||   |||   |||   |||:::|||   :::||
     389 AAAATACAGGTTATTCCAAGGCGGATGAAGTGTTTGAATTTT...ATCAC 435

     108 rAlaLysAla 111
         |   ||||||
     436 AATCAAGGCT 445

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAC85071

seq_documentation_block:
ID   AAC85071 standard; DNA; 868 BP.
XX
AC   AAC85071;
XX
DT   08-MAY-2001  (first entry)
XX
DE   Atherosclerosis-associated gene seq ID No. 7.
XX
KW   Atherosclerosis-associated gene; stroke; myocardial infarction; human;
KW   ischemia; coronary artery disease; angina pectoris; hypertension;
KW   peripheral vascular disease; renal artery stenosis; antiatherosclerotic;
KW   cerebroprotective; cardiant; gene therapy; hypotensive; vasotropic;
KW   antianginal; ds.
XX
OS   Homo sapiens.
XX
PN   WO200104264-A2.
XX
PD   18-JAN-2001.
XX
PF   28-JUN-2000; 2000WO-US17887.
XX
PR   07-JUL-1999;   99US-0349015.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Jones KA,  Volkmuth W,  Walker MG;
XX
DR   WPI; 2001-138330/14.
XX
PT   Composition comprising atherosclerosis-associated polynucleotide useful
PT   in diagnosis, prognosis, treatment, and prevention of atherosclerosis
PT   and stroke, myocardial infarction, or hypertension  -
XX
PS   Claim 1; Page 39; 58pp; English.
XX
CC   The invention provides novel atherosclerosis-associated polynucleotides
CC   and polypeptides encoded by the genes. Expression vectors and host cells
CC   for producing the polypeptides are disclosed and methods for screening or
CC   purifying ligands which specifically bind to the polypeptides are also
CC   provided. The polynucleotides are useful for treating diseases associated
CC   with the altered expression of a gene that is coexpressed with one or
CC   more known atherosclerosis-associated genes in a subject. They are
CC   useful in diagnosis, prognosis, treatment, prevention, selection and
CC   evaluation of therapies for atherosclerosis including stroke, myocardial
CC   infarction, transient cerebral ischemia, mesenteric ischemia, coronary
CC   artery disease, angina pectoris, peripheral vascular disease, renal
CC   artery stenosis, and hypertension. Sequences AAC85065-85098 represent
CC   atherosclerosis-associated genes of the invention.
XX
SQ   Sequence 868 BP; 249 A; 207 C; 182 G; 230 T; 0 other;


alignment_scores:
            Quality:   69.50              Length:     80
              Ratio:  1.418                Gaps:      2
 Percent Similarity:  61.250   Percent Identity: 25.000

alignment_block:
 US-10-048-197-2 x AAC85071   ..

 Align seg 1/1  to: AAC85071  from: 1  to: 868

      22 CysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIl 38
         |||||||||   ::::::::::::   ::::::   :::   :::
     196 TGTGTTTCCTACTCTTCTTTGTCTTGGAGGAAAGCAAAGGGAGAGAAATG 245

      38 eGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerA 55
            :::   ::::::   |||      |||:::||||||::::::||||
     246 GACTTCACCAGTGGTCTTTGGCTTCATCAATTCACAGGAAATGGCATCAA 295

      55 rgPheGlySerAlaAspSerIleSerPheMetIleValValIleLysPhe 71
         ||:::   :::   |||   |||:::      |||::::::|||   |||
     296 GATGGTTCAACTAAGACATGATCACTAAAAACATTATAATAATACCTTTT 345

      72 GlyHisThrAlaIleLeuAlaProAsnArgTrp.............GlnGl 84
            |||   :::|||   .   :::::::::              :::
     346 TGAAAAACTCAGTTTCTCCTGTTTACTAAATATTTATTTCATCAACATGG 395

      84 uIleLeuSerLeuIleIleSerPhe...LeuTrpValLys 96
         :::   |||  .:::   :::|||   :::|||||:::
     396 GCTGCGTTCCACTGTGTCAGGATTCTGCATGTGGGTGGAG 435

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC43179

seq_documentation_block:
ID   AAC43179 standard; DNA; 1014 BP.
XX
AC   AAC43179;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 38318.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
```

KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.

```
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

alignment_scores:
            Quality:   69.50              Length:    155
              Ratio:  0.952                Gaps:      8
 Percent Similarity:  47.097   Percent Identity: 23.226

alignment_block:
 US-10-048-197-2 x AAC43179   ..

 Align seg 1/1  to: AAC43179  from: 1  to: 1014

       3 AsnHisHisIleArgLeuThrIleSerAlaLeuLeuThr..... 15
         |||:::::::::   :::|||:::||||||      |||
     181 AACAGAAGAGTATCGGTTACCGTTTCCGCCGCCGCAACAACTGAACCTCT 230

      16 .AlaLeuLeuValThrGlyCysValSer...ThrGlyAsnValAla.... 29
          ::::::||||||||||||||:::   :::   ||||||:::::::::
     231 CACCGTCCTCGTTACCGGTGCCGGTGGAAGAACAGGACAAATTGTGTATA 280

      30 .....MetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLys 44
           :::||||||:::::::::|||   :::   :::   :::::::::
     281 AGAAATTGAAGGAGAGGTCGGAGCAGTTCGTGGCAAGGGGTTTAGTGAGG 330

      45 GlyLysThrAsnLysGlnGluIleSerSerArgPheGlySerAlaAspSe 61
           |||   :::|||::::::|||:::::::::            |||::
     331 ACAAAGGAGAGTAAGGAGAAGATCAATGGAGAG.....GATGA 368

      61 rIleSerPheMetIleValValIleLysPheGlyHisThrAlaIleLeuA 78

          :           :::   :::      :::      ||||||   :::|
     369 A.....GTATTCATTGGAGATATTAGAGATACTGCATCAATTG 406

      78 laProAsnArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrp 94
         |||||        :::   |||   :::|||:::|||
     407 CTCCT...GCTGTTGAAGGGATTGATGCTTTGGTCATTCTTACTAGCGCT 453

      95 ValLysProTyrArgPro.............................. 100
         |||         :::|||
     454 GTTCCGCAGATGAAGCCCGGTTTTGATCCTAGTAAAGGAGGAAGACCTGA 503

     100 ..... 100

     504 GTTCTTCTTTGATGATGGAGCTTATCCAGAACAGGTTGACTGGATTGGTC 553

     101 ..LysAsn.............................LeuSerPheTyr 106
           ||||||                            |||||||||
     554 AGAAGAATCAGATAGACGCTGGTGTTGATGATGATGGATTGTCGTTTGTT 603

     107 LeuThrAlaLysAla 111
         :::||||||||||||
     604 GTAACAGCTAAGGCT 618

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:AAT42063

seq_documentation_block:
ID   AAT42063 standard; DNA; 1830121 BP.
XX
AC   AAT42063;
XX
DT   14-SEP-1999  (first entry)
XX
DE   Haemophilus influenzae complete genome sequence.
XX
KW   Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW   expression modulating fragment; regulation; gene expression; vector;
KW   organism; open reading frame; ORF; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO9633276-A1.
XX
PD   24-OCT-1996.
XX
PF   22-APR-1996;   96WO-US05320.
XX
PR   07-JUN-1995;   95US-0487429.
PR   21-APR-1995;   95US-0426787.
PR   07-JUN-1995;   95US-0476102.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Adams MD,  Fleischmann RD,  Smith HO,  Venter JC,  White O;
XX
DR   WPI; 1996-485782/48.
XX
PT   Haemophilus influenzae Rd genome recorded on computer readable
PT   medium  - useful for identifying commercially important nucleic acid
PT   fragments by homology searching
XX
PS   Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC   This sequence represents the complete genome sequence of the bacterium
CC   Haemophilus influenzae strain Rd.  The invention relates to a computer
CC   readable medium (CRM) having recorded upon it the complete H.influenzae
CC   nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC   sequence at least 99% identical to (I).  By providing the full-length
CC   genomic sequence in a computer readable form, it is possible to identify
CC   commercially important nucleic acid fragments and expression modulating
CC   fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC   regulate the expression of a nucleic acid molecule.  Vectors and altered
CC   organisms comprising the predicted ORFS can be used to produce any of the
```

```
CC   polypeptide fragments of the H. influenzae Rd genome.
XX
SQ   Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;


alignment_scores:
            Quality:   69.50              Length:     70
              Ratio:  1.655                Gaps:      2
 Percent Similarity:  60.000   Percent Identity: 27.143

alignment_block:
 US-10-048-197-2 x AAT42063   ..

 Align seg 1/1  to: AAT42063  from: 1  to: 1830121

       2 LeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLe 18
         |||:::::|||:::   :::   :::   :::|||   :::   |||||
  259928 TTATCTCGCCATTTTATAGTCTATTTATTTAGCTTATGTGCGATCTTACT 259977

      18 uValThrGlyCys.....................ValSerThrGlyAsnV 28
         |::::::||||||                     ||||||   ::::::|
  259978 CCTTGCTGGCTGTGTTCAGTCTAGGGGCGGTTTTGTGTCTAAAAATCATG 260027

      28 alAlaMetLysGluGlnAsnGlnGlnThr..................... 37
         ||:::::::   |||||||||   :::|||
  260028 TCGTGTTGGCTGAACAAAACCCGAATACTCACTTTGAACAAGAAGTGATG 260077

      38 .........IleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGl 51
                 :::::::|||:::::::::   ||||||   ::::::::::||
  260078 ATCGTGCGATTAAGCCAAGTACTACTGGTTGGAAAAATGAGTAATGAAGA 260127

      51 uIleSerSer 54
         |   :::|||
  260128 ACGCGCATCG 260137

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAD12599

seq_documentation_block:
ID   AAD12599 standard; cDNA; 1155 BP.
XX
AC   AAD12599;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human protein having hydrophobic domain encoding cDNA clone HP10790.
XX
KW   Human; hydrophobic domain; gene therapy; nutritional supplement;
KW   cell proliferation; immunomodulatory; autoimmune disorder; antimicrobial;
KW   multiple sclerosis; rheumatoid arthritis; insulin-dependent diabetes;
KW   haematopoiesis; tissue growth activity; Parkinson's disease; cytostatic;
KW   Huntington's disease; Alzheimer's disease; chemotactic; chemokinetic;
KW   haemostatic; thrombolytic; tumour growth inhibitor; anabolic;
KW   contraceptive; antiinfertility; antiinflammatory; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             110..1102
FT                   /*tag= a
FT                   /product= "Human protein having hydrophobic domain"
FT                   /note= "CDS is specifically is claimed in claim 3"
XX
PN   WO200149728-A2.
XX
PD   12-JUL-2001.
XX
PF   28-DEC-2000; 2000WO-JP09359.
XX
PR   06-JAN-2000; 2000JP-0000585.
PR   06-JAN-2000; 2000JP-0000588.
PR   11-JAN-2000; 2000JP-0002299.
PR   03-FEB-2000; 2000JP-0026862.
PR   03-MAR-2000; 2000JP-0058367.
XX
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENT.
XX
PI   Kato S,  Kimura T;
XX
DR   WPI; 2001-418355/44.
DR   P-PSDB; AAE06604.
XX
PT   Human proteins with hydrophobic domains and the nucleic acids encoding
PT   them, useful for preventing diagnosing and treating e.g. cancer,
PT   Alzheimer's and inflammation -
XX
PS   Claim 4; Page 469-471; 563pp; English.
XX
CC   The present sequence is human protein with hydrophobic domain encoding
CC   cDNA clone HP10790. The polynucleotide and polypeptide of the invention
CC   may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate polypeptide expression. The polynucleotides
CC   may be used to produce the polypeptide, by inserting the nucleic acids
CC   into a host cell and culturing the cell to express the protein. The
CC   polynucleotides and its complementary sequences may also be used as DNA
CC   probes in diagnostic assays and also used in gene therapy. The
CC   polypeptides may also be used as antigens in the production of antibodies
CC   and in assays to identify modulators of polypeptide expression and
CC   activity. The polypeptides and nucleic acids may be used as nutritional
CC   supplements, to modulate cytokine and cell proliferation activity, to
CC   modulate immune stimulation or suppression (e.g. for the treatment of
CC   microbial infections and autoimmune disorders such as multiple sclerosis,
CC   rheumatoid arthritis and insulin-dependent diabetes), to modulate
CC   haematopoiesis, to modulate tissue growth activity (e.g. for the
CC   treatment of Parkinson's disease, Huntington's disease and Alzheimer's
CC   disease), to modulate activin and inhibin activity (e.g. for controlling
CC   fertility), to modulate chemotactic and chemokinetic activity, to
CC   modulate haemostatic and thrombolytic activity, to modulate receptor
CC   ligand activity, to modulate inflammation and to inhibit tumour growth.
XX
SQ   Sequence 1155 BP; 280 A; 265 C; 274 G; 336 T; 0 other;


alignment_scores:
            Quality:   69.00              Length:     73
              Ratio:  1.643                Gaps:      4
 Percent Similarity:  57.534   Percent Identity: 32.877

alignment_block:
 US-10-048-197-2 x AAD12599/rev   ..

 Align seg 1/1  to reverse of: AAD12599  from: 1  to: 1155

       5 HisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGl 21
         |||::::::::::|||:::|||:::|||:::      |||      |||||
     891 CATGTCCAGGTCACAGTATCCAGCTTGTTTGGGTATCTGAGGAGGACTGG 842

      21 yCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrI 38
         ||||      ||||||:::   :::||||||      :::   ::::::|
     841 CTGC......ACTGGAACTCCTGGAATGAAGGCTCCTGGTTTAAAAGTAA 798

      38 leGluGln............ThrIleIleLysGlyLysThrAsnLysGln 50
         ||:::|||             :::   :::|||||||||
     797 TCAAACAGGAACGATTAGTACAAGTACCTTCTGGGAAAACT......... 757

      51 GluIleSerSerArgPheGlySerAla......AspSerIleSerPheMe 65
                 |||   :::|||   :::      |||   :::   |||::
     756 .........AGTATCTGGGGCCATTCTCCTCCTGATGTTGTTCGCTTTAT 716

      65 tIleValValIleLysPhe 71
         :|||      |||   |||
     715 TATTTCATTTATTGTGTTT 697

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA81456

seq_documentation_block:
```

```
ID   AAA81456 standard; DNA; 46593 BP.
XX
AC   AAA81456;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_4 SEQ ID NO:4.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 260-274; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 46593 BP; 11355 A; 13195 C; 11355 G; 10687 T; 1 other;


alignment_scores:
            Quality:   69.00              Length:     89
              Ratio:  1.278                Gaps:      1
 Percent Similarity:  60.674   Percent Identity: 28.090

alignment_block:
 US-10-048-197-2 x AAA81456/rev   ..

 Align seg 1/1  to reverse of: AAA81456  from: 1  to: 46593

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||:::   :::||||||   ::::::||||||   ::::::||||||
   46127 CGCATCATCGTTTCGGCTGCGTTCGTTGCGTTGGCATTAGCAGGTTGCGG 46078

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
          |||   :::||||||:::::::::::::   :::   |||
   46077 CTCAATCAATAATGTAACCGTTTCCGACTAGAAACTTCAGGAACGTGCCG 46028

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
                :::      :::   |||      :::||||||:::|||
   46027 CGTTTGCCTTGGGCGTCAGCCCAAATGCCGTAAAAATCAGCAACCGCAAC 45978

      57 GlySerAlaAspSerIleSerPheMetIleValVal.IleLysPheGlyH 73
         :::::::::      |||:::|||      :::||| ::::::
   45977 AATGAAGGCATACGCATCAACTTTACCGCAACTGTGGGTAAGCGCGTGAG 45928

      73 isThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIle 89
           :::|||:::|||                |||:::   |||   |||
   45927 CCAATGCTATGTTACCAGTGTAATCAGCACAATCGGCGTTACCACTTCCG 45878

      90 IleSerPheLeuTrp 94
         ::::::|||:::|||
   45877 ATGCAATTTGTTTGG 45863

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAF21612

seq_documentation_block:
ID   AAF21612 standard; DNA; 349980 BP.
XX
AC   AAF21612;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW   diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW   ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200066791-A1.
XX
PD   09-NOV-2000.
XX
PF   08-MAR-2000; 2000WO-US05928.
XX
PR   30-APR-1999;   99US-0132068.
PR   08-OCT-1999;   99WO-US23573.
PR   28-FEB-2000; 2000GB-0004695.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Pizza M,  Hickey E,  Peterson J,  Tettelin H,  Venter JC,  Masignani V;
PI   Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V,  Rappuoli R;
PI   Frazer CM,  Grandi G;
XX
DR   WPI; 2000-647603/62.
XX
PT   Neisseria meningitidis B full length genome sequence and open reading
PT   frames are used to detect, treat and prevent Neisserial infections -
XX
PS   Claim 7; Appendix A; 692pp; English.
XX
CC   The present invention describes the full length genome of
CC   Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC   to AAF21613 represent fragments of the NMB genomic sequence, as the
CC   sequence was too long to go in a record on its own it was split into 8
CC   sequences which overlap each other at the beginning and end of each
```

```
CC   sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC   the beginning  of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC   the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC   Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC   AAF21606 represent PCR primers which are used in the exemplification of
CC   the present invention. The NMB genome and fragments from it have
CC   antibacterial activity, and can be used in vaccines and gene therapy.
CC   Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC   proteins can be used in compositions for treating or preventing infection
CC   due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC   presence of Neisserial bacteria or of antibodies raised to Neisserial
CC   bacteria. Computers, computer memory, computer storage medium or computer
CC   databases can be used in a search to identify open reading frames (ORFs)
CC   or coding sequences within the NMB genome. The DNA sequences provide
CC   further opportunities to find antigenic or immunogenic proteins which are
CC   more effective in vaccines than the outer membrane proteins currently
CC   used.
XX
SQ   Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;


alignment_scores:
           Quality:   69.00              Length:     89
             Ratio:  1.278                Gaps:      1
 Percent Similarity:  60.674   Percent Identity: 28.090

alignment_block:
 US-10-048-197-2 x AAF21612/rev   ..

 Align seg 1/1  to reverse of: AAF21612  from: 1  to: 349980

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||:::   :::||||||   ::::::||||||   ::::::||||||
  239247 CGCATCATCGTTTCGGCTGCGTTCGTTGCGTTGGCATTAGCAGGTTGCGG 239198

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
          |||   :::||||||:::::::::::::   :::   |||
  239197 CTCAATCAATAATGTAACCGTTTCCGACTAGAAACTTCAGGAACGTGCCG 239148

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
                :::      :::   |||      :::|||||||:::|||
  239147 CGTTTGCCTTGGGCGTCAGCCCAAATGCCGTAAAAATCAGCAACCGCAAC 239098

      57 GlySerAlaAspSerIleSerPheMetIleValVal.IleLysPheGlyH 73
         :::::::::      |||:::|||      :::||| ::::::
  239097 AATGAAGGCATACGCATCAACTTTACCGCAACTGTGGGTAAGCGCGTGAG 239048

      73 isThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIle 89
         :::|||:::|||                |||:::   |||   |||
  239047 CCAATGCTATGTTACCAGTGTAATCAGCACAATCGGCGTTACCACTTCCG 238998

      90 IleSerPheLeuTrp 94
         :::::::|||:::|||
  238997 ATGCAATTTGTTTGG 238983

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA81489

seq_documentation_block:
ID   AAA81489 standard; DNA; 837096 BP.
XX
AC   AAA81489;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_37 SEQ ID NO:37.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 629-865; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC   represent specifically claimed Neisseria meningitidis genomic DNA
CC   sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC   Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC   AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC   isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC   AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC   sequences, which are all used in the exemplification of the present
CC   invention. The nucleic acid sequences, protein sequences, and antibodies
CC   against them, can be used in the manufacture of a composition. The
CC   composition can be used as a medicament (or in the manufacture of a
CC   medicament) for treating, preventing or diagnosing infection due to
CC   Neisserial bacteria. For example, some of the identified proteins could
CC   be components of vaccines against Meningococcus B; against all serotypes;
CC   and/or against all pathogenic Neissariae. Identification of sequences
CC   from the bacterium will also facilitate production of biological probes,
CC   particularly organism-specific probes. Attempts to make efficacious
CC   Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC   Multivalent vaccines have also been tried but none have successfully
CC   overcome antigenic variability. The provision of further, complete
CC   sequences may provide an opportunity to identify secreted or surface
CC   exposed proteins that may be presumed targets for the immune system and
CC   which are not antigenically variable or at least more conserved than
CC   other more variable regions.
XX
SQ   Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;


alignment_scores:
           Quality:   69.00              Length:     89
             Ratio:  1.278                Gaps:      1
 Percent Similarity:  60.674   Percent Identity: 28.090

alignment_block:
 US-10-048-197-2 x AAA81489/rev   ..

 Align seg 1/1  to reverse of: AAA81489  from: 1  to: 837096

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||:::   :::||||||   ::::::||||||   ::::::||||||
  604018 CGCATCATCGTTTCGGCTGCGTTCGTTGCGTTGGCATTAGCAGGTTGCGG 603969

      23 lSerThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluG 40
          |||   :::||||||:::::::::::::   :::   |||
  603968 CTCAATCAATAATGTAACCGTTTCCGACTAGAAACTTCAGGAACGTGCCG 603919

      40 lnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPhe 56
                :::      :::   |||      :::|||||||:::|||
  603918 CGTTTGCCTTGGGCGTCAGCCCAAATGCCGTAAAAATCAGCAACCGCAAC 603869
```

```
     57 GlySerAlaAspSerIleSerPheMetIleValVal.IleLysPheGlyH 73
        :::::::::      |||:::|||      :::||| ::::::
 603868 AATGAAGGCATACGCATCAACTTTACCGCAACTGTGGGTAAGCGCGTGAG 603819

     73 isThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIle 89
          :::|||:::|||                |||:::   |||   |||
 603818 CCAATGCTATGTTACCAGTGTAATCAGCACAATCGGCGTTACCACTTCCG 603769

     90 IleSerPheLeuTrp 94
        ::::::|||:::|||
603768 ATGCAATTTGTTTGG 603754
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABA03041

seq_documentation_block:

ID ABA03041 standard; DNA; 2944528 BP.
XX
AC ABA03041;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes EGD-e genome sequence.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease; ds.
XX
OS Listeria monocytogenes.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR01118.
XX
PR 11-APR-2000; 2000FR-0004629.
XX
PA (INSP) INST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
DR WPI; 2002-010914/01.
XX
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -
XX
PS Claim 1; SEQ ID No 1; 192pp; French.
XX
CC The present sequence is the genome sequence of Listeria monocytogenes
CC EGD-e. This sequence and fragments of this sequence are useful for
CC selecting probes and primers for detecting genes in L. monocytogenes and
CC related organisms, and to study genetic polymorphisms and other genomes.
CC Proteins (ABB47297-ABB50149) expressed from the present sequence are
CC useful for raising specific antibodies, identification of L.
CC monocytogenes and related organisms, and for biosynthesis and
CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
CC proteins encoded by it are also useful for selecting compounds that
CC regulate gene expression and cell replication and modulate L.
CC monocytogenes-related diseases. In addition, this sequence and proteins
CC encoded by it are useful in pharmaceutical and vaccines compositions for
CC the treatment or prevention of infections by L. monocytogenes and related
CC organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

alignment_scores:
Quality: 68.50 Length: 49
Ratio: 1.851 Gaps: 1
Percent Similarity: 75.510 Percent Identity: 34.694

alignment_block:
US-10-048-197-2 x ABA03041/rev ..

Align seg 1/1 to reverse of: ABA03041 from: 1 to: 2944528

```
      4 HisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValTh 20
        ||||||   :::      |||:::::::::||||||::::::|||:::
1863702CACCATGCGCATGCGTTAATAAAATCCGTTTTAACGGTGATTTTAATTTA 1863653

     20 rGlyCysValSer...ThrGlyAsnValAlaMetLysGluGlnAsnGlnG 36
        |||   |||:::         :::   |||:::|||:::|||   |||:
1863652TGGATAAGTGAAATGATTCCTTTCGCATGCGCTAAAATCCCAGTATCAAT 1863603

     36 lnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGlu 51
        ::|||:::::::::   ::::::|||   ::::::|||:::||||||
1863602CAACGTTAAGCCATCTCTTTCTAAAACCAAGTAACAATTCACAGGAA 1863556
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC37886

seq_documentation_block:

ID AAC37886 standard; DNA; 470 BP.
XX
AC AAC37886;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 19019.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.

```
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

alignment_scores:
            Quality:   68.00              Length:     59
              Ratio:  1.789                Gaps:      3
 Percent Similarity:  64.407   Percent Identity: 35.593

alignment_block:
 US-10-048-197-2 x AAC37886/rev   ..

 Align seg 1/1  to reverse of: AAC37886  from: 1  to: 470

      46 LysThrAsnLysGlnGluIleSerSerArgPheGlySerAlaAspSerIl 62
         ||||||   |||   ::::::||||||         |||::::::|||::
     347 AAAACCCTAAAAGCTTCCTTGAGCTCT.........TCCTCTGAATCAGT 307

      62 eSerPheMetIleValValIleLysPhe.........GlyHisThrAlaI 76
         :||||||:::::::::::::|||:::|||        |||   :::   :
     306 GTCCTTCATTTTCCTTGCCATCAGGTTCAGGAACTCCGGGAAATCAATGG 257

      76 leLeuAlaProAsnArgTrpGlnGluIleLeuSer..LeuIleIleSerP 92
         ::      |||:::        :::::::::     |||   :::||||
     256 TCCCGTTCCCATCTGCATCAACTTCATTGATCATGTCTTGAAGCTCAGCT 207

      92 heLeuTrpValLysProTyrArgPro 100
         ||   ||||||   |||   ||||||
     206 TCTGTTGGGTTTTGCCCCAGAGACCT 181

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC99183

seq_documentation_block:
ID   AAC99183 standard; cDNA; 584 BP.
XX
AC   AAC99183;
XX
DT   09-MAR-2001  (first entry)
XX
DE   Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:411.
XX
KW   Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW   detection; diagnosis; identification; cytostatic; neuroprotective;
KW   nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW   antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW   linkage analysis; tissue identification; tissue typing; forensic;
KW   neural; immune system; muscular; reproductive; gastrointestinal;
KW   pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055320-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05989.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-579444/54.
DR   P-PSDB; AAB54418.
XX
PT   New nucleic acid that is a pancreatic cancer antigen for preventing,
PT   treating, or ameliorating a medical condition, particular pancreatic
PT   cancer, or for use in assays for diagnosing a pathological condition -
XX
PS   Claim 1; Page 845; 1379pp; English.
XX
CC   AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC   proteins, called pancreatic cancer antigens, given in AAB54008 to
CC   AAB54466. The human pancreatic cancer antigens have cytostatic,
CC   neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
CC   gynaecological, cardiant and antiinflammatory activities, and can be used
CC   in gene therapy. The polynucleotide and proteins can be used for
CC   preventing, treating, or ameliorating a medical condition or in assays
CC   for diagnosing a pathological condition or a susceptibility to one in a
CC   subject. Binding partners to the proteins and the activity of the
CC   proteins can be identified. The pancreatic cancer antigens can be used to
CC   detect, treat or prevent pancreatic disorders, especially cancer.
CC   Agonists and antagonists to the antigens can be screened for. The
CC   pancreatic cancer antigen polynucleotides can be used to design nucleic
CC   acid hybridisation probes that can be used in chromosome mapping, linkage
CC   analysis, tissue identification and/or typing and a variety of forensic
CC   and diagnostic methods. The proteins can be used to generate antibodies
CC   which are used to purify, detect and target the polypeptides, including
CC   both in vivo and in vitro diagnostic and therapeutic methods. The
CC   proteins can be used to treat or prevent neural, immune system, muscular,
CC   reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC   proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 584 BP; 165 A; 99 C; 164 G; 148 T; 8 other;


alignment_scores:
            Quality:   68.00              Length:     84
              Ratio:  1.333                Gaps:      4
 Percent Similarity:  60.714   Percent Identity: 30.952

alignment_block:
 US-10-048-197-2 x AAC99183   ..

 Align seg 1/1  to: AAC99183  from: 1  to: 584

       8 LeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSe 24
         :::::::::   ||||||::::::|||::::::|||:::      |||::
     242 GTGGTCGTGGTGGCTTTGGTGGCAGCCGTGGTGGTGGTGGATATGGTGGC 291

      24 rThrGly.............AsnValAlaMetLysGlu.GlnAsnGlnGln 36
         ::::|||             :::::::::|||   ||| |||      :::
     292 AGTGGGGATGGCTATAATGGATTTGGTAATGATGGAAGCAATTTTGGGAA 341

      37 ThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSe 53
         :::::::|||   |||:::|||   |||:::   |||:::|||      ||
     342 GTGGTGGAAGCTACAATGATTTTGGGGAATTACAACAATCAG......TC 385

      53 rSerArgPheGlySerAlaAspSerIleSerPheMetIleValValIleL 70
         ||||:::||||||:::       :::   :::|||
     386 TTCAAATTTTGGGACCATGAAGGGAAGAAATTTT.............. 419

      70 ysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeu 86
            :::      ||||||      |||||||:::   ||||||
     420 ........GGAAGCAGAACTCTGGCCCTATGGCGGTGGAAGCCAATACTT 461

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC35309

seq_documentation_block:
ID   AAC35309 standard; DNA; 1176 BP.
XX
AC   AAC35309;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 9720.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
```

KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.

```
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

alignment_scores:
            Quality:   68.00              Length:     59
              Ratio:  1.619                Gaps:      3
 Percent Similarity:  71.186   Percent Identity: 28.814

alignment_block:
 US-10-048-197-2 x AAC35309    ..

 Align seg 1/1  to: AAC35309  from: 1  to: 1176

       3 AsnHisHisIleArgLeuThrIleSerAlaLeuLeuThr..... 15
         |||:::::::::   :::|||:::||||||     |||
     263 AACAGAAGAGTATCGGTTACCGTTTCCGCCGCCGCAACAACTGAACCTCT 312

      16 .AlaLeuLeuValThrGlyCysValSer...ThrGlyAsnValAla.... 29
         ::::::||||||||||||||:::   :::   ||||||:::::::::
     313 CACCGTCCTCGTTACCGGTGCCGGTGGAAGAACAGGACAAATTGTGTATA 362

      30 .....MetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleIleLys 44
           :::||||||:::::::::|||   :::   :::   :::::::::
     363 AGAAATTGAAGGAGAGGTCGGAGCAGTTCGTGGCAAGGGGTTTAGTGAGG 412

      45 GlyLysThrAsnLysGlnGluIleSer 53
         |||   :::|||:::::::|||:::
     413 ACAAAGGAGAGTAAGGAGAAGATCAAT 439

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAD21310

seq_documentation_block:
ID   AAD21310 standard; cDNA; 2144 BP.
XX
AC   AAD21310;
XX
DT   28-JAN-2002  (first entry)
XX
DE   Arabidopsis thaliana SDS cDNA.
XX
KW   Meiotic prophase I; chromosome 1; male sterile SDS mutant; apomixis;
KW   plant breeding; seed production; SDS protein; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             98..1834
FT                   /*tag=  a
FT                   /product= "SDS protein"
XX
PN   WO200174144-A1.
XX
PD   11-OCT-2001.
XX
PF   29-MAR-2001; 2001WO-US09875.
XX
PR   31-MAR-2000; 2000US-193523P.
XX
PA   (UYPE-) UNIV PENNSYLVANIA STATE.
XX
PI   Ma H;
XX
DR   WPI; 2001-662939/76.
DR   P-PSDB; AAE12998.
XX
PT   A new plant gene from Arabidopsis, designated SDS, mutations in which
PT   are associated with inability to produce pollen, is useful for the
PT   production of male sterile plants for plant breeding  -
XX
PS   Claim 8; Page 39; 47pp; English.
XX
CC   The patent discloses novel plant gene from Arabidopsis thaliana,
CC   designated SDS, which is associated with a failure to maintain
CC   homologue attachment during meiotic prophase I. The SDS gene is
CC   located on chromosome 1 and is useful in plant breeding to produce
CC   male sterile SDS mutants and cloned progeny by apomixis where
CC   meiosis is bypassed in seed production. The gene may also be
CC   used as a probe to identify related genes in other plant species,
CC   and to identify and isolate other genes of the meiosis regulatory
CC   pathway. The present sequence is a cDNA encoding Arabidopsis thaliana
CC   SDS protein.
XX
SQ   Sequence 2144 BP; 649 A; 461 C; 467 G; 567 T; 0 other;

alignment_scores:
            Quality:   68.00              Length:    123
              Ratio:  0.958                Gaps:      9
 Percent Similarity:  57.724   Percent Identity: 31.707

alignment_block:
 US-10-048-197-2 x AAD21310    ..

 Align seg 1/1  to: AAD21310  from: 1  to: 2144

       5 HisIleArgLeuThr.IleSerAlaLeuLeuThrAlaLeuLeuVal...T 20
         |||:::|||   ||| |||:::::::   |||:::||||||   |||   |
    1766 CATGTTAGAACAACAGATAACGAGTTGCCTGAATGCGTTAAGAGTCTGGA 1815

      20 hrGlyCysVal.SerThrGlyAsnValAlaMetLysGluGln........ 33
         ||||||||::: :::::::::|||      :::|||||||||
    1816 CTGGTTGCTTGGGCAGTAAGCAAT......CAAAAAGAACAAAAACCCTA 1859
```

```
      34 .AsnGlnGlnThrIle..............GluGlnThrIleIleLysG 45
         ||||||:::|||:::                 |||:::||||||
    1860 AAACCAGGACACAGTATACTCCGATACCAACACACAGGTTATCATTACTA 1909

      45 lyLysThrAsnLysGlnGluIleSerSerArgPheGlySerAlaAsp... 60
            ||||||::::::::::|||::::::   :::|||:::|||
    1910 TTTACAAAAACAAACACAAGGTAAGTAATAAGAACTCCTCTACAGATTTA 1959

      61 ..............................SerIleSerPheMetIl 66
                                       ||||||   ::::::
    1960 TATACTTAATCGAGCTGGACTTAATTAGCTCTTAGTATACCAATTATTAG 2009

      66 eValValIleLysPheGlyHisThr......AlaIleLeuAlaProAsnA 81
         ::::::|||      :::||||||        ::::::   ||||||
    2010 TGCCACCATTTGTGTCGCTCATACACATTTATTTCTTATTTTCCCTAATT 2059

      81 rgTrp...GlnGluIleLeuSerLeuIleIleSerPheLeuTrpValLys 96
            ::::::||||||:::      ||||||::::::      |||
    2060 CATTAGACTCTCATATTCTTAAAAAGAATATTTCCTTGTTTGAAAAAAAA 2109

      97 ProTyrArgProLys 101
            :::   |||
    2110 AAAAAAAAAAAAAA 2124
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH17234

seq_documentation_block:

ID AAH17234 standard; cDNA; 2303 BP.
XX
AC AAH17234;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:16615.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 16615; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2303 BP; 703 A; 566 C; 502 G; 532 T; 0 other;

alignment_scores:
Quality: 68.00 Length: 124
Ratio: 1.062 Gaps: 6
Percent Similarity: 51.613 Percent Identity: 26.613

alignment_block:
US-10-048-197-2 x AAH17234/rev ..

Align seg 1/1 to reverse of: AAH17234 from: 1 to: 2303

```
       2 LeuAsnHisHisIle.ArgLeuThrIleSerAlaLeuLeuThrAlaLeuL 18
         |||||||||:::::: ||||||:::      :::::::|||:::::|||
    1151 CTTAATCATCAGGTGCAGGCTTTCTGCGCTGTCCATACTCTCTACGCTTT 1102

      18 euValThrGlyCys..............................ValSer 24
         :::   ||||||                                 |||
    1101 CTACAATGGGCTGCGGCCTTGGTCGGTCAGCCTTCCTCACTGGGGTAGTC 1052

      25 ThrGlyAsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluGlnTh 41
         |||:::      :::   |||:::   ::::::::::::::::::
    1051 TTGGGGTCCTCAGAGTTGTTGGAGTCTGTGTCAGACTCACTCAGTGACCT 1002

      41 rIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPheGlyS 58
         |||   :::      :::      ::::::||||||   |||:::|
    1001 CTGCATC...AGCTGTCTCAGCCTGGCTAACTTCAGTTCCCTCTTCTCCA 955

      58 erAla..................................................... 59
         ||:::
     954 GTGGGATCTTTTTAGAATTTCTAGAGGAAGCCTGAGCATCCTGGAATTCC 905

      60 ...AspSerIleSerPheMetIleValValIleLysPheGlyHisThrAl 75
         ||||||:::|||::::::   :::                   :::
     904 AAGGACAGTTTTTCCTGGGTGCAGACA..............TTTTCTGA 870

      75 aIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIleSerP 92
         |||      ||||||:::|||::::::|||||||||         ||||
     869 GATTTCCTTTCCCAATAACTGGCGTAAGATTTTATCAGAATGTTCATCTT 820

      92 he.LeuTrpValLysPro 97
         || |||||||||   |||
     819 TTGCTTTGGGTCTAGCCC 802
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV52205

seq_documentation_block:

ID AAV52205 standard; DNA; 14872 BP.
XX
AC AAV52205;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:72.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.

```
XX
OS   Streptococcus pneumoniae.
XX
PN   WO9818931-A2.
XX
PD   07-MAY-1998.
XX
PF   30-OCT-1997;   97WO-US19588.
XX
PR   31-OCT-1996;   96US-0029960.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Choi GH,  Dillon PJ,  Dougherty BA,  Fannon M;
PI   Kunsch CA,  Rosen CA;
XX
DR   WPI; 1998-272225/24.
XX
PT   Computer-readable medium with recorded Streptococcus pneumoniae
PT   polynucleotide sequences - useful in diagnostic kits and assays, and
PT   pharmaceutical compositions and vaccines for Streptococcus
PT   pneumoniae
XX
PS   Claim 1; Page 593-602; 1409pp; English.
XX
CC   The present invention describes a computer readable medium which has
CC   the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC   recorded on it, or a representative fragment or a sequence at least 95%
CC   identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC   SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC   Streptococcus pneumoniae. The present invention also describes an
CC   isolated nucleic acid molecule encoding a homologue of any of the
CC   fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC   nucleic acid molecule is produced by a process comprising: (a) screening
CC   a genomic DNA library using as a probe a target sequence defined by any
CC   of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC   library which contain sequences that hybridise to the target sequence and
CC   isolating the nucleic acid molecules from the members; or (b) isolating
CC   mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC   molecules whose nucleotide sequence is homologous to amplification
CC   primers derived from the fragment of the S. pneumoniae genome to prime
CC   the amplification and isolating the amplified sequences. The computer
CC   readable medium can be used in a computer-based system for identifying
CC   fragments of the S. pneumoniae genome of commercial importance, or
CC   expression modulating fragments of the S. pneumoniae genome. Products
CC   from the present invention can be used in diagnosis kits and assays, and
CC   pharmaceutical compositions and vaccines for S. pneumoniae.
XX
SQ   Sequence 14872 BP; 4340 A; 2874 C; 3210 G; 4448 T; 0 other;

alignment_scores:
            Quality:   68.00              Length:    121
              Ratio:  1.000                Gaps:      5
 Percent Similarity:  56.198   Percent Identity: 22.314

alignment_block:
 US-10-048-197-2 x AAV52205   ..

 Align seg 1/1  to: AAV52205  from: 1  to: 14872

       5 HisIleArgLeuThrIleSerAlaLeuLeuThrAla..............  16
         |||   :::|||||||   :::::::|||   :::|||
    8446 CACTGCAAACTTACACGTGGGAGCCTTTATGTTGCCCAAGTATGTTGAGG 8495

      17 .....................................LeuLeuValThrG  21
                                              |||||||||   |
    8496 ACATTTTAGAAGAAGAGGAAGGAAAAAAATGAGTCGTTTACTAGTTATTG 8545

      21 lyCysValSerThrGlyAsnValAlaMetLysGlu......GlnAsnGln  35
         |||||   ::::::::::::|||||::::::::      |||::::::
    8546 GTTGTGGGGGCGTTGCCCAAGTTGCTATTTCAAAGATTTGTCAAGATAGC 8595

      36 GlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln..... 50
         :::|||:::   :::   :::|||   :::::::|||:::::::::
    8596 GAAACATTTACAGAGATTATGATTGCTAGCCGTACCAAGTCAAAATGCGA 8645

      51 .GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleV 67
         :::::::::::::::::   :::      |||   ::::::
    8646 TGACTTGAAAGCGAAGCTAGAAGGCAAAACAAGTACTAAAATT....... 8688

      67 alValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGln 83
                         :::||||||   |||      ::::::   :::
    8689 ................GAAACTGCAGCACTTGATGCTGACAAGGTTGAA 8721

      84 GluIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPr 100
         |||:::::::::|||||   |||:::         ||||||
    8722 GAAGTGATTGCCCTGATTGAAAGCTAC.........AAACCAGAAGCTGT 8762

     100 oLysAsnLeuSer 104
             |||::::::
    8763 TTTGAATGTAGCT 8775

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH15139

seq_documentation_block:
ID   AAH15139 standard; cDNA; 1055 BP.
XX
AC   AAH15139;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:13197.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 13197; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
```

```
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1055 BP; 280 A; 215 C; 211 G; 349 T; 0 other;


alignment_scores:
            Quality:   67.50              Length:    111
              Ratio:  1.350                Gaps:      4
 Percent Similarity:  45.045   Percent Identity: 27.027

alignment_block:
 US-10-048-197-2 x AAH15139   ..

 Align seg 1/1  to: AAH15139  from: 1  to: 1055

      27 AsnValAlaMetLysGluGlnAsnGlnGlnThrIleGluGlnThrIleIl 43
         ||||||:::         :::|||:::         :::   :::|||
     607 AACGTGACAGCTGCGTGGAGAAATAATGGTTGGCAACACTTTTCTATTGA 656

      43 eLysGlyLysThrAsnLysGlnGluIleSerSerArgPheGlySerAlaA 60
          ||||||   ::::::||||||:::||||||   ||||||
     657 GAAGGGATTTAATTCCAAGCAGAATATTAGCATCCGTTTT.......... 696

      60 spSerIleSerPheMetIleValValIleLysPheGlyHisThrAlaIle 76
                           :::||||||||||||:::
     697 ....................CTTGTAATTAAGTTCAACTTT......TGT 720

      77 LeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIleIleSerPheLe 93
         |||:::   :::   |||:::|||:::   :::|||   :::|||   ||
     721 TTATCTTCGGAGGGTTGGAAAGAGTTAAGTGAGTTAGGACTTAGTCATTT 770

      93 u.............................TrpValLysProTyr.... 98
         |                            |||:::|||:::
     771 GGATGCATTCCTTTAAATACAGTGCATTACTAAAGTCAGACCTTTTTGAA 820

      98 ................................................. 98

     821 ATAACTTTTTAAAGTCTGTACCGAGTCCTGTTTAACGTAAAATAGCTAGA 870

      99 ...ArgProLysAsnLeuSerPheTyrLeuThr 108
            ||||||:::|||:::|||      ||||||
     871 GAAAGGCCAAACAACATTTCCTCCTCCCTTACT 903

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ51314

seq_documentation_block:
ID   AAZ51314 standard; DNA; 1555 BP.
XX
AC   AAZ51314;
XX
DT   06-JUN-2000  (first entry)
XX
DE   S. cerevisiae delta-9 desaturase gene - p1-ole1.
XX
KW   Yeast delta-9 desaturase; p1-ole1 gene; yeast; transgenic plant;
KW   oil plant; mono-unsaturated fatty acid; foodstuff; ss.
XX
OS   Saccharomyces cerevisiae.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             11..1553
FT                   /*tag= a
FT                   /transl_except= (pos: 920..922, aa: Leu)
FT                   /product= "delta-9 desaturase"
XX
PN   WO200011012-A1.
XX
PD   02-MAR-2000.
XX
PF   24-AUG-1999;   99WO-US19443.
XX
PR   24-AUG-1998;   98US-0097586.
XX
PA   (RUTF ) UNIV RUTGERS STATE NEW JERSEY.
XX
PI   Martin CE,  Mitchell A;
XX
DR   WPI; 2000-237610/20.
DR   P-PSDB; AAY70269.
XX
PT   New synthetic desaturase gene, useful to obtain transgenic plants that
PT   produce a higher yield of unsaturated fatty acids, is customized for
PT   expression in a plant cytoplasm  -
XX
PS   Claim 6; Page 71; 77pp; English.
XX
CC   The present sequence is modified yeast delta-9 desaturase p1-ole1 gene
CC   comprising a desaturase domain and a cyt b5 domain, and customised for
CC   expression in a plant cytoplasm. This gene is modified for expression in
CC   Arabidopsis and related species to obtain transgenic plants that produce
CC   a higher yield of unsaturated fatty acids. Transgenic plants, especially
CC   oil plants, are produced by this method which have increased amounts of
CC   unsaturated, particularly mono-unsaturated fatty acids, to obtain
CC   foodstuff that is more healthy and has improved flavour.
XX
SQ   Sequence 1555 BP; 424 A; 336 C; 333 G; 462 T; 0 other;


alignment_scores:
            Quality:   67.50              Length:    104
              Ratio:  1.055                Gaps:      1
 Percent Similarity:  61.538   Percent Identity: 16.346

alignment_block:
 US-10-048-197-2 x AAZ51314   ..

 Align seg 1/1  to: AAZ51314  from: 1  to: 1555

       7 ArgLeuThrIleSerAlaLeuLeuThrAlaLeuLeuValThrGlyCysVa 23
         |||   ||||||:::::::::::::::::::|||||:::         ::
    1005 AGGGATACCATAACTTCCATCATGAGTTCCCTACTGATTATAGAAACGCT 1054

      23 lSerThrGlyAsn..................................... 27
         :|||:::|||:::
    1055 ATCAAGTGGTACCAATACGATCCTACTAAAGTGATCATCTACTTGACTTC 1104

      28 ..............ValAlaMetLysGluGlnAsnGlnGlnThrIleGlu 39
                        :::::::::::::::::::::::   :::|||::::::
    1105 TCTCGTGGGACTTGCTTACGATCTCAAGAAGTTCTCTCAAAACGCTATCG 1154

      40 GlnThrIleIleLysGlyLysThrAsnLysGlnGluIleSerSerArgPh 56
         :::   :::::::::::::|||:::::::::::::::          |||::
    1155 AGGAGGCTCTTATCCAACAAGAGCAAAAGAAGATCAACAAGAAGAAGGCT 1204

      56 eGlySerAlaAspSerIleSerPheMetIleValValIleLysPheGlyH 73
         :          ::::::   :::::::|||   :::       ::::::
    1205 AAGATTAATTGGGGACCTGTTCTTACTGATCTTCCTATGTGGGATAAGCA 1254

      73 isThrAlaIleLeuAlaProAsnArgTrpGlnGluIleLeuSerLeuIle 89
           :::::::|||:::   :::|||   ::::::::::||||||   :::
    1255 AACTTTCCTTGCTAAGTCTAAGGAGAACAAGGGACTTGTTATCATCTCTG 1304

      90 IleSerPheLeu 93
           ||||||:::
    1305 GAATCGTTCATG 1316
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:AAT27729

seq_documentation_block:
ID   AAT27729 standard; DNA; 2100 BP.
XX
AC   AAT27729;
XX
DT   17-AUG-1996  (first entry)
XX
DE   Opacity associated protein gene.
XX
KW   Opacity associated protein; OapA; OapB; vector; vaccine; antibody;
KW   diagnosis; immunogen; ds.
XX
OS   Haemophilus influenzae strain Rd variant H175.
XX
FH   Key             Location/Qualifiers
FT   CDS             209..1495
FT                   /*tag=  a
FT                   /product= OapA
FT   CDS             1552..2008
FT                   /*tag=  b
FT                   /product= OapB
XX
PN   WO9613515-A1.
XX
PD   09-MAY-1996.
XX
PF   24-OCT-1995;   95WO-US13672.
XX
PR   31-OCT-1994;   94US-0332576.
XX
PA   (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.
XX
PI   Weiser JN;
XX
DR   WPI; 1996-239445/24.
DR   P-PSDB; AAR95909;
DR   P-PSDB; AAR95910.
XX
PT   New opacity associated protein of Haemophilus influenzae - and
PT   related DNA, vectors, transformed hosts and antibodies, useful in
PT   vaccines and for diagnosis.
XX
PS   Claim 4; Fig 1a-c; 48pp; English.
XX
CC   A DNA fragment (T27729) of Haemophilus influenzae includes
CC   open reading frames coding for opacity associated proteins OapA
CC   (R95909) and OapB (R95910).  It was isolated from the spontaneous
CC   opacity variant H175 by cloning Sau3AI fragments in lambda EMBL,
CC   infecting E, coli, and screening clones for transforming activity
CC   of the opaque type.  The oap gene can be used for prodn. of
CC   recombinant Oap proteins useful as immunogens for vaccines.  Viral
CC   vectors including the gene may be developed as live viral vaccines.
CC   Detection of H. influenzae in a sample can be performed using the
CC   gene as a probe.
XX
SQ   Sequence 2100 BP; 735 A; 396 C; 412 G; 557 T; 0 other;


alignment_scores:
            Quality:   67.50              Length:     42
              Ratio:  2.045                Gaps:      1
 Percent Similarity:  78.571   Percent Identity: 33.333

alignment_block:
 US-10-048-197-2 x AAT27729   ..

 Align seg 1/1  to: AAT27729  from: 1  to: 2100

      11 SerAlaLeuLeuThrAlaLeuLeuValThrGlyCysValSerThrGlyAs 27
         |||   ::::::||||||||||||||:::||||||||||||||:::::::::||
    1568 TCTCTTATTTTTACCGCACTTTTAATGACTGGCTGTGTGCAAAATGCGAA 1617
```

```
      27 nValAlaMetLysGluGlnAsnGlnGlnThrIleGlu.........GlnT 41
         ||||:::       :::   :::::::::   :::|||         ::::
    1618 TGTAACAACACCTCAAGCGCAAAAAATGCAAGTAGAAAAAGTGGATAAAG 1667

      41 hrIleIleLysGlyLysThrAsnLys 49
         :::::   ||||||:::::::::::::
    1668 CCTTACAAAAAGGCGAAGCTGATCGA 1693

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAF18124

seq_documentation_block:
ID   AAF18124 standard; DNA; 3474 BP.
XX
AC   AAF18124;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polynucleotide sequence SEQ ID 143.
XX
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliferative disorder; wound healing; infectious disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05918.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   P-PSDB; AAB58248.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 1; Page 608-609; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence 3474 BP; 1062 A; 692 C; 689 G; 1015 T; 16 other;


alignment_scores:
```

```
           Quality:   67.50             Length:    100
             Ratio:  1.298               Gaps:      4
 Percent Similarity:  52.000   Percent Identity: 28.000

alignment_block:
 US-10-048-197-2 x AAF18124/rev   ..

 Align seg 1/1  to reverse of: AAF18124  from: 1  to: 3474

      19 ValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGl 35
         |||||||||      :::|||   |||         |||:::|||:::||
     697 GTGACAGGGACGGAAGACACCAAGAACAGCAACAAAAAGCAGCAGCAGCA 648

      35 nGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluI 52
         ||||      :::               ||||||::::::   |||   :
     647 GCAGCCGGGCAGGTGG............GGAAAGGCTGAGGCACAATGTG 610

      52 leSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValVal 68
         ::|||   :::|||      :::   ::::::|||   :::   :::|||
     609 TCAGTATCCATTTCCACATTGTACTGAAACTCTCGGCACTGGAAATTGTC 560

      69 IleLysPheGly.HisThrAlaIleLeuAlaPro.............. 79
            :::   ||| ||||||:::   :::   |||
     559 TCATCAGAAGGAACATACTGGAAGGGTAAGACCCGATGCAAGGCAAGGAG 510

      80 .....AsnArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrp 94
              :::||||||              |||:::::::|||   ||||||
     509 AGAGCTCGAGGTGGTGG............CTGCTGGTTAGCTCCCTCTGG 472

      95 ValLysProTyrArgProLysAsnLeuSerPheTyrLeuThrAlaLys 110
         ::::::   :::         |||            :::|||   :::
     471 GCGCGGATCTGGCCAGAAGTGAATGCACCAACATCCATAACTCCTCGG 424

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV19980

seq_documentation_block:
ID   AAV19980 standard; cDNA; 4177 BP.
XX
AC   AAV19980;
XX
DT   15-JUL-1998  (first entry)
XX
DE   Full length Keratinocyte growth factor-2 coding sequence.
XX
KW   Keratinocyte growth factor-2; KGF-2; fibroblast growth factor-12; FGF-12;
KW   keratinocyte proliferation; growth stimulator; skin strength; mucositis;
KW   epidermal thickening; wound healing; inflammatory bowel disease; therapy;
KW   inflammation; hair growth promoter; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             593..1219
FT                   /*tag= a
XX
PN   WO9806844-A1.
XX
PD   19-FEB-1998.
XX
PF   13-AUG-1997;   97WO-US14112.
XX
PR   28-FEB-1997;   97US-0039045.
PR   13-AUG-1996;   96US-0023852.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Coleman TA,  Duan R,  Jimenez P,  Mendrick D,  Moore PA;
PI   Ni J,  Rampy MA,  Ruben SM,  Zhang J;
XX
DR   WPI; 1998-159536/14.
DR   P-PSDB; AAW53793.
XX
PT   Keratinocyte growth factor-2 deletion mutants - useful to promote
PT   or accelerate wound healing
XX
PS   Disclosure; Fig 3; 251pp; English.
XX
CC   This sequence encodes the human keratinocyte growth factor-2 (KGF-2). The
CC   protein is used to create the mutants of the invention, that stimulate
CC   proliferation of keratinocytes. The mutants have enhanced keratinocyte
CC   growth stimulating activity as compared to wild-type KGF-2 (also known as
CC   fibroblast growth factor-12). They are used to stimulate growth or
CC   proliferation of keratinocytes. In particular, they are used to prevent
CC   or improve the appearance of wrinkles or aged skin, improving skin
CC   strength, promoting epidermal thickening, reducing scarring or improving
CC   healing after cosmetic surgery. The mutants are also useful for promoting
CC   wound healing, especially where an individual is wound healing impaired.
CC   Wounds to be treated may be surgical or excisional wounds, deep wounds
CC   involving damage of the dermis and epidermis, eye tissue wounds, dental
CC   tissue wounds, oral cavity wounds, diabetic, dermal, cubitus, arterial or
CC   venous stasis ulcers or burns. Treatment, especially by anastomosis, of
CC   wounds caused by colonic or gastrointestinal surgical procedures can also
CC   be achieved through use of the KGF-2 deletion mutants. They can also be
CC   used for treatment or prevention of mucositis, inflammatory bowel
CC   disease, reduction of inflammation, promoting hair growth, urothelial
CC   healing or tissue growth or repair in the female genital tract, or for
CC   treating tissue exposed to radiation or protecting tissue to be exposed
CC   to radiation (e.g. to allow an increase in radiation dosage used to treat
CC   a malignancy).
XX
SQ   Sequence 4177 BP; 1332 A; 816 C; 874 G; 1155 T; 0 other;


alignment_scores:
           Quality:   67.50             Length:    100
             Ratio:  1.298               Gaps:      4
 Percent Similarity:  52.000   Percent Identity: 28.000

alignment_block:
 US-10-048-197-2 x AAV19980/rev   ..

 Align seg 1/1  to reverse of: AAV19980  from: 1  to: 4177

      19 ValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGl 35
         |||||||||      :::|||   |||         |||:::|||:::||
     699 GTGACAGGGACGGAAGACACCAAGAACAGCAACAAAAAGCAGCAGCAGCA 650

      35 nGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluI 52
         ||||      :::               ||||||::::::   |||   :
     649 GCAGCCGGGCAGGTGG............GGAAAGGCTGAGGCACAATGTG 612

      52 leSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValVal 68
         ::|||   :::|||      :::   :::::::|||   :::   :::|||
     611 TCAGTATCCATTTCCACATTGTACTGAAACTCTCGGCACTGGAAATTGTC 562

      69 IleLysPheGly.HisThrAlaIleLeuAlaPro.............. 79
            :::   ||| ||||||:::   :::   |||
     561 TCATCAGAAGGAACATACTGGAAGGGTAAGACCCGATGCAAGGCAAGGAG 512

      80 .....AsnArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrp 94
              :::||||||              |||:::::::|||   ||||||
     511 AGAGCTCGAGGTGGTGG............CTGCTGGTTAGCTCCCTCTGG 474

      95 ValLysProTyrArgProLysAsnLeuSerPheTyrLeuThrAlaLys 110
         ::::::   :::         |||            :::|||   :::
     473 GCGCGGATCTGGCCAGAAGTGAATGCACCAACATCCATAACTCCTCGG 426

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ11169

seq_documentation_block:
ID   AAZ11169 standard; DNA; 4177 BP.
XX
AC   AAZ11169;
XX
DT   03-NOV-1999  (first entry)
XX
```

```
DE   Human keratinocyte growth factor, KGF-2, gene.
XX
KW   Keratinocyte growth factor; KGF-2; human; platelet; fibrinogen; albumin;
KW   globulin; total serum protein; blood; hypofibrinogenaemia; cirrhosis;
KW   disseminated intravascular coagulation; thrombocytopaenia; myelofibrosis;
KW   hypoalbuminaemia; posttransfusion purpura; metastatic tumour; anaemia;
KW   leukaemia; haemolytic syndrome; Zieve's syndrome; rheumatic disease;
KW   HELLP preeclamptic syndrome; congenital rubella syndrome; systemic lupus;
KW   Epstein-Barr infectious mononucleosis; thyrotoxicosis; uraemia; therapy;
KW   infection; tissue necrosis; vasculitis; ulcerative bowel disease;
KW   serositis; subacute bacterial endocarditis; liver disease; amyloidosis;
KW   congestive heart failure; constrictive pericarditis; nephrotic syndrome;
KW   cardiac valvular disease; hypoglobulinaemia; keratoconjuctivitis sicca;
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             593..1216
FT                   /*tag= a
FT                   /product= KGF-2
XX
PN   WO9941282-A1.
XX
PD   19-AUG-1999.
XX
PF   12-FEB-1999;   99WO-US03018.
XX
PR   30-DEC-1998;   98US-0114387.
PR   13-FEB-1998;   98US-0074585.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Jimenez P,  Louie A,  Mendrick D,  Rampy MA,  Russell D;
XX
DR   WPI; 1999-527359/44.
DR   P-PSDB; AAY32915.
XX
PT   Use of keratinocyte growth factor-2 to increase levels of platelets,
PT   fibrinogen, albumin, globulin and total serum protein
XX
PS   Disclosure; Fig 3; 331pp; English.
XX
CC   This sequence encodes the human keratinocyte growth factor-2 (KGF-2)
CC   protein. Fragments and mutants of this sequence are used in the methods
CC   of the invention, for increasing the level of platelets, fibrinogen,
CC   albumin, globulin, and total serum protein in the blood. KGF-2 can also
CC   be used to stimulate proliferation of salivary gland cells, lacrimal
CC   gland cells, sinus epithelium, and Goblet cells. The methods can also be
CC   used to treat hypofibrinogenaemia caused by a cirrhosis, and disseminated
CC   intravascular coagulation (DIC). The methods can be used to treat
CC   thrombocytopaenia and to alleviate hypoalbuminaemia. These diseases are
CC   caused by: drug induced hypersensitivity, thrombocytopaenia purpura,
CC   posttransfusion purpura, metastatic tumours in the bone, aplastic
CC   anaemia, myelofibrosis, leukaemia, haemolytic syndromes, cancer
CC   chemotherapy, Zieve's syndrome, sepsis, HELLP preeclamptic syndrome,
CC   megaloblastic anaemia peritonitis, congenital rubella syndrome,
CC   Epstein-Barr infectious mononucleosis, systemic lupus, preeclampsia,
CC   thyrotoxicosis, uraemia, rheumatic diseases, granulomatous processes,
CC   bacterial viral and parasitic infections, tissue necrosis, vasculitis,
CC   ulcerative bowel disease, serositis, subacute bacterial endocarditis,
CC   liver disease, amyloidosis, malnutrition, malignancy, congestive heart
CC   failure, constrictive pericarditis, cardiac valvular disease, nephrotic
CC   syndrome, trauma and crush injuries, gastrointestinal and lymphatic
CC   fistulae, and protein-losing gastroenteropathies. The methods can also be
CC   used to treat hypoglobulinaemia, total protein loss, damage to the sinus
CC   epithelium, and can be used to increase proliferation of epithelial cells
CC   of the bladder or prostate, stimulate proliferation of the salivary gland
CC   cells and to increase Goblet cell proliferation for treating or
CC   preventing keratoconjuctivitis sicca.
XX
SQ   Sequence 4177 BP; 1332 A; 816 C; 874 G; 1155 T; 0 other;
```

```
alignment_scores:
            Quality:   67.50              Length:    100
              Ratio:  1.298                Gaps:      4
 Percent Similarity:  52.000   Percent Identity: 28.000

alignment_block:
 US-10-048-197-2 x AAZ11169/rev   ..

 Align seq 1/1  to reverse of: AAZ11169  from: 1  to: 4177

      19 ValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGl 35
         |||||||||      :::|||   |||          |||:::|||:::||
     699 GTGACAGGGACGGAAGACACCAAGAACAGCAACAAAAAGCAGCAGCAGCA 650

      35 nGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluI 52
         ||||      :::               ||||||::::::   |||   :
     649 GCAGCCGGGCAGGTGG.............GGAAAGGCTGAGGCACAATGTG 612

      52 leSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValVal 68
         ::|||   :::|||      :::   ::::::|||   :::   :::|||
     611 TCAGTATCCATTTCCACATTGTACTGAAACTCTCGGCACTGGAAATTGTC 562

      69 IleLysPheGly.HisThrAlaIleLeuAlaPro................. 79
            :::   ||| ||||||:::   :::   |||
     561 TCATCAGAAGGAACATACTGGAAGGGTAAGACCCGATGCAAGGCAAGGAG 512

      80 .....AsnArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrp 94
           :::||||||              |||:::::|||   ||||||
     511 AGAGCTCGAGGTGGTGG.............CTGCTGGTTAGCTCCCTCTGG 474

      95 ValLysProTyrArgProLysAsnLeuSerPheTyrLeuThrAlaLys 110
         ::::::   :::         |||          :::|||   :::
     473 GCGCGGATCTGGCCAGAAGTGAATGCACCAACATCCATAACTCCTCGG 426

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA71214

seq_documentation_block:
ID   AAA71214 standard; DNA; 4177 BP.
XX
AC   AAA71214;
XX
DT   20-NOV-2000  (first entry)
XX
DE   Human keratinocyte growth factor KGF-2 genomic DNA SEQ ID NO: 23.
XX
KW   Human; keratinocyte growth factor; KGF-2; antiulcer; antidiabetic;
KW   antiinflammatory; cytoprotective; dermatological; gastrointestinal;
KW   hepatic; respiratory; renal; cerebroprotective; mucositis; treatment;
KW   epithelial cell proliferation; inflammatory bowel disease; lung damage;
KW   liver disorder; diabetes; oral injury; gastrointestinal injury;
KW   gut toxicity; gastric; duodenal; epidermolysis bullosa; skin graft;
KW   skin disorder; renal failure; brain injury; intestinal fibrosis;
KW   proctitis, female reproductive tract disorder; pulmonary fibrosis;
KW   pneumonitis; pleural retraction; hemopoietic syndrome; myelotoxicity; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             593..1219
FT                   /*tag= a
FT                   /product= "KGF-2"
XX
PN   US6077692-A.
XX
PD   20-JUN-2000.
XX
PF   13-FEB-1998;   98US-0023082.
XX
PR   13-AUG-1996;   96US-0023852.
PR   28-FEB-1997;   97US-0039045.
PR   13-AUG-1997;   97US-0055561.
PR   05-JUN-1995;   95US-0461195.
```

```
PR   14-FEB-1995;   95WO-US01790.
PR   23-MAY-1997;   97US-0862432.
PR   13-AUG-1997;   97US-0910875.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Mendrick D,  Duan DR,  Ni J,  Jimenez P,  Coleman TA,  Gruber JR;
PI   Dillon PJ,  Gentz RL,  Ruben SM,  Zhang J,  Moore PA,  Rampy MA;
XX
DR   WPI; 2000-441307/38.
DR   P-PSDB; AAB10304.
XX
PT   Novel keratinocyte growth factor useful for promoting and accelerating
PT   wound healing, comprising at least 10 contiguous amino acids from a
PT   specific amino acid sequence -
XX
PS   Disclosure; Fig 3A-D; 190pp; English.
XX
CC   This invention describes a novel human keratinocyte growth factor, KGF-2
CC   (I), which has antiulcer, antidiabetic, antiinflammatory, cytoprotective,
CC   dermatological, gastrointestinal, hepatic, respiratory, renal and
CC   cerebroprotective activity. (I) is useful for stimulating epithelial cell
CC   proliferation in patients suffering from wound, mucositis, ulcer such as
CC   venous statis ulcer, diabetic ulcer and cubitus ulcer. (I) is also useful
CC   for treating inflammatory bowel disease, liver disorder, lung damage,
CC   diabetes, oral injury, gastrointestinal injury, gut toxicity, gastric
CC   ulcer, duodenal ulcer, epidermolysis bullosa, skin graft, skin disorder,
CC   renal failure, brain injury, breast tissue injury, urothelial damage,
CC   female reproductive tract disorder, intestinal fibrosis, proctitis,
CC   pulmonary fibrosis, pneumonitis, pleural retraction, hemopoietic syndrome
CC   and myelotoxicity. (I) is also useful for increasing the adherence of
CC   skin grafts to wound beds and to stimulate re-epithelialization from the
CC   wound bed, to produce changes in hepatocyte proliferation, to reduce the
CC   side effects of gut toxicity, to regenerate skin in full and partial
CC   thickness skin defects, and to prevent and heal damage to lungs. KGF-2
CC   shows enhanced activity, increased stability, higher yield and better
CC   solubility. This sequence encodes the human KGF-2 protein described in
CC   the method of the invention.
XX
SQ   Sequence 4177 BP; 1332 A; 816 C; 874 G; 1155 T; 0 other;


alignment_scores:
            Quality:   67.50              Length:    100
              Ratio:  1.298                Gaps:      4
 Percent Similarity:  52.000   Percent Identity: 28.000

alignment_block:
 US-10-048-197-2 x AAA71214/rev   ..

 Align seg 1/1  to reverse of: AAA71214  from: 1  to: 4177

      19 ValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGl 35
         |||||||||      :::|||   |||         |||:::|||:::||
     699 GTGACAGGGACGGAAGACACCAAGAACAGCAACAAAAAGCAGCAGCAGCA 650

      35 nGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluI 52
         ||||      :::              ||||||::::::   |||   :
     649 GCAGCCGGGCAGGTGG.............GGAAAGGCTGAGGCACAATGTG 612

      52 leSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValVal 68
         ::|||   :::|||      :::   ::::::|||   :::   :::|||
     611 TCAGTATCCATTTCGACATTGTACTGAAACTCTCGGCACTGGAAATTGTC 562

      69 IleLysPheGly.HisThrAlaIleLeuAlaPro.................. 79
            :::   ||| ||||||:::   :::   |||
     561 TCATCAGAAGGAACATACTGGAAGGGTAAGACCCGATGCAAGGCAAGGAG 512

      80 .....AsnArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrp 94
              :::||||||            |||:::::::|||   ||||||
     511 AGAGCTCGAGGTGGTGG.............CTGCTGGTTAGCTCCCTCTGG 474

      95 ValLysProTyrArgProLysAsnLeuSerPheTyrLeuThrAlaLys 110
         ::::::   :::         |||            :::|||   :::
     473 GCGCGGATCTGGCCAGAAGTGAATGCACCAACATCCATAACTCCTCGG 426

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF31912

seq_documentation_block:
ID   AAF31912 standard; DNA; 4177 BP.
XX
AC   AAF31912;
XX
DT   10-APR-2001  (first entry)
XX
DE   Human keratinocyte growth factor coding sequence #2.
XX
KW   Keratinocyte growth factor; KGF-2; epithelial cell proliferation; wound;
KW   mucositis; ulcer; inflammatory bowel disease; liver disorder;
KW   lung damage; diabetes; oral injury; gastrointestinal injury;
KW   epidermolysis bullosa; renal failure; brain injury; proctitis;
KW   pulmonary fibrosis; haemopoietic syndrome; ovary injury; infertility;
KW   liver fibrosis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200102433-A1.
XX
PD   11-JAN-2001.
XX
PF   03-JUL-2000; 2000WO-US18328.
XX
PR   02-JUL-1999;   99US-0142343.
PR   14-JUL-1999;   99US-0143648.
PR   15-JUL-1999;   99US-0144024.
PR   12-AUG-1999;   99US-0148628.
PR   19-AUG-1999;   99US-0149935.
PR   03-NOV-1999;   99US-0163375.
PR   22-DEC-1999;   99US-0171677.
PR   19-APR-2000; 2000US-0198322.
PR   19-MAY-2000; 2000US-0205417.
PR   30-JUN-2000; 2000US-9911224.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Jimenez P,  Duan DR,  Rampy MA,  Mendrick D,  Zhang J;
PI   Ni J,  Moore PA,  Coleman TA,  Gruber JR,  Dillon PJ,  Gentz RL;
XX
DR   WPI; 2001-071578/08.
DR   P-PSDB; AAB61668.
XX
PT   A polynucleotide encoding the human keratinocyte growth factor useful
PT   for stimulating epithelial cell proliferation in a patients that has
PT   e.g a wound -
XX
PS   Disclosure; Fig 3; 591pp; English.
XX
CC   The present sequence is a coding sequence for human keratinocyte growth
CC   factor (KGF-2). KGF-2 can be used to stimulate epithelial cell
CC   proliferation in a patient, where the patient has a wound, mucositis, an
CC   ulcer, inflammatory bowel disease, liver disorder, lung damage, diabetes,
CC   oral injury, gastrointestinal injury, gut toxicity, epidermolysis
CC   bullosa, skin graft, skin disorder, renal failure, brain injury, breast
CC   tissue injury, urothelial damage, female reproductive tract disorder,
CC   intestinal fibrosis, proctitis, pulmonary fibrosis, peunrumonitis, plural
CC   retraction, haemopoietic syndrome, and myelotoxicity. In addition, KGF-2
CC   can be used in the treatment or prevention of ovary injury, infertility,
CC   or fibrosis of the liver. KGF-2 also promotes internal healing, donor
CC   site healing, internal surgical wound healing or healing of incisional
CC   wounds made during cosmetic surgery in a patient.
XX
SQ   Sequence 4177 BP; 1332 A; 816 C; 874 G; 1155 T; 0 other;


alignment_scores:
            Quality:   67.50              Length:    100
```

```
              Ratio:  1.298                Gaps:      4
 Percent Similarity:  52.000   Percent Identity: 28.000

alignment_block:
 US-10-048-197-2 x AAF31912/rev   ..

 Align seg 1/1  to reverse of: AAF31912  from: 1  to: 4177

      19 ValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnAsnGl 35
         |||||||||      :::|||   |||        |||:::|||:::||
     699 GTGACAGGGACGGAAGACACCAAGAACAGCAACAAAAAGCAGCAGCAGCA 650

      35 nGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGlnGluI 52
         ||||      :::              ||||||::::::   |||   :
     649 GCAGCCGGGCAGGTGG...........GGAAAGGCTGAGGCACAATGTG 612

      52 leSerSerArgPheGlySerAlaAspSerIleSerPheMetIleValVal 68
         ::|||   :::|||      :::   :::::::|||   :::   :::|||
     611 TCAGTATCCATTTCCACATTGTACTGAAACTCTCGGCACTGGAAATTGTC 562

      69 IleLysPheGly.HisThrAlaIleLeuAlaPro..............  79
            :::   ||| ||||||:::   :::   |||
     561 TCATCAGAAGGAACATACTGGAAGGGTAAGACCCGATGCAAGGCAAGGAG 512

      80 .....AsnArgTrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrp 94
              :::|||||||              |||:::::|||   ||||||
     511 AGAGCTCGAGGTGGTGG...........CTGCTGGTTAGCTCCCTCTGG 474

      95 ValLysProTyrArgProLysAsnLeuSerPheTyrLeuThrAlaLys 110
         ::::::   :::         |||            :::|||   :::
     473 GCGCGGATCTGGCCAGAAGTGAATGCACCAACATCCATAACTCCTCGG 426

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL26466

seq_documentation_block:
ID   ABL26466 standard; DNA; 13239 BP.
XX
AC   ABL26466;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 30871.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 30871; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 13239 BP; 3611 A; 2940 C; 3072 G; 3616 T; 0 other;


alignment_scores:
           Quality:  67.50              Length:    114
             Ratio:  1.107                Gaps:      4
 Percent Similarity:  53.509   Percent Identity: 21.930

alignment_block:
 US-10-048-197-2 x ABL26466/rev   ..

 Align seg 1/1  to reverse of: ABL26466  from: 1  to: 13239

       1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
         |||||||:::   |||       :::             |||      :::||
    7963 ATGCTCGGGGCACATTGTACAATA.....CTACACCCAACATT 7926

      17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnA 34
         |      |||         |||:::   :::   |||   ::::::::::
    7925 ACATTGCACACATGGGCCAAGTGTGCGTGAGTGTGCGTCGAGATCTCGAT 7876

      34 snGlnGlnThrIleGluGlnThrIleIleLys.................. 44
         :::::   |||:::||||||         :::
    7875 CTCGTTATACTGTAGAACAACTATATAGAGAATCGGGCATTGCCCAAGCG 7826

      45 .........GlyLysThrAsnLysGlnGluIleSer.............. 53
                 |||:::   ::::::   |||   :::
    7825 AACCGAACCGGGAGAGATCGGCGAGCGGAGTGGGCGCTTATCAGGAATCA 7776

      54 .......SerArgPheGlySerAlaAspSerIleSerPheMetIle.... 66
               :::   :::||||||:::::::|||:::::::|||   :::
    7775 TCCAATTCAAGTTTACGGGTCAGGCAATTCGGTCCAATTCAATCTGCCGT 7726

      67 .....ValValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArg 81
               :::   ::::::   |||   |||||||:::   |||   |||:::
    7725 GATTTACCCACCTCCGCAGAGGAATCACCGCCGTAACTGCCATGAATCAT 7676

      82 TrpGlnGluIleLeuSerLeuIleIleSerPheLeuTrpVal 95
         |||:::::|||:::::::::   :::   :::   |||:::
    7675 TGGAACTCAATCATAACAATATACGTTCTAATTCATTGGTTA 7634
```

THIS PAGE BLANK (USPTO)

OM of: US-10-048-197-2 to: GenEmbl:* out_format : pfs

Date: Sep 18, 2002 9:39 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10048197/runat_17092002_135956_23539/app_query.fasta_1.168
-DB=GenEmbl -QFMT=fastap -SUFFIX=oli.rge -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10048197_@CGN1_1_6377 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-10-048-197-2
Query length: 111
Database: GenEmbl:*
Database sequences: 1797656
Database length: 1873333701
Search time (sec): 1904.340000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation ..
gb_pat:AX081233	+	111.00	2095.80	2.3e-108	336	! AX081233 Sequence 1 from Patent
gb_pat:AX067463	+	111.00	2056.95	3.3e-106	94750	! AX067463 Sequence 38 from Pat
gb_htg:AC099920	-	9.00	125.05	133.11	48309	! AC099920 Mus musculus clone R
gb_htg:AC103410	+	9.00	123.09	171.19	64235	! AC103410 Mus musculus clone R
gb_pr:AC064841	+	9.00	122.51	184.30	69829	! AC064841 Homo sapiens chromos
gb_htg:AL670662	+	9.00	119.15	283.64	113776	! AL670662 Homo sapiens chromo
gb_pr:AF241734	+	9.00	118.79	297.12	119919	! AF241734 Homo sapiens PAC M1
gb_htg:AC103115	-	9.00	116.72	387.36	161924	! AC103115 Rattus norvegicus c
em_htg_hum:AC019286	+	9.00	116.67	389.92	163134	! Ac019286 Homo sapiens clone
gb_pr:AC108881	+	9.00	116.67	390.12	163230	! AC108881 Homo sapiens chromo
em_htg_hum:AC055844	+	9.00	116.63	391.78	164017	! Ac055844 Homo sapiens chromo
gb_htg:AC010811	-	9.00	116.57	394.72	165412	! AC010811 Homo sapiens chromo
gb_htg:AC106051	+	9.00	116.44	401.38	168577	! AC106051 Homo sapiens chromo
gb_pr:AC097661	-	9.00	116.42	402.69	169197	! AC097661 Homo sapiens chromo
gb_ro:AC074041	-	9.00	116.36	405.53	170548	! AC074041 Mus musculus chromo
gb_htg:AC099145	+	9.00	116.30	408.94	172176	! AC099145 Rattus norvegicus c
gb_htg:AC013798	+	9.00	116.16	416.34	175706	! AC013798 Homo sapiens clone
gb_htg:AC009290	-	9.00	116.04	422.54	178670	! AC009290 Homo sapiens clone
gb_ro:AC074224	-	9.00	115.66	444.02	188989	! AC074224 Mus musculus chromo
gb_htg:AC023140	-	9.00	115.46	455.59	194575	! AC023140 Homo sapiens chromo
gb_htg:AC068980	+	9.00	115.10	477.06	204991	! AC068980 Homo sapiens chromo
gb_htg:AC024447	+	9.00	115.10	477.07	204993	! AC024447 Homo sapiens chromo
gb_htg:AC092942	+	9.00	115.07	478.77	205825	! AC092942 Homo sapiens chromo
gb_htg:AC099581	+	9.00	114.95	486.35	209518	! AC099581 Mus musculus chromo
gb_ba:AP003590	-	9.00	111.75	733.23	333500	! AP003590 Nostoc sp. PCC 7120
gb_pat:AX120599	+	8.00	142.02	15.11	261	! AX120599 Sequence 515 from Pate
gb_pl:AF314550	+	8.00	141.22	16.73	293	! AF314550 Glycine max clone NTN5
gb_pl:AF222879	+	8.00	140.79	17.69	312	! AF222879 Glycine max clone R14
gb_in:TIU42193	-	8.00	138.96	22.37	407	! U42193 Taeniothrips inconsequen
gb_pl:AF060192	+	8.00	137.28	27.72	519	! AF060192 Glycine max putative r
gb_pl:AB027824	+	8.00	137.26	27.82	521	! AB027824 Schizosaccharomyces po
gb_pat:AX317918	-	8.00	136.73	29.74	562	! AX317918 Sequence 19 from Paten
gb_pl:AY063021	+	8.00	133.91	42.73	847	! AY063021 Arabidopsis thaliana p
gb_in:HAY12271	+	8.00	133.75	43.62	867	! Y12271 H.armigera mRNA for puta
gb_in:HAY12269	+	8.00	133.73	43.75	870	! Y12269 H.armigera mRNA for puta
gb_pat:AR020714	-	8.00	133.61	44.42	885	! AR020714 Sequence 3 from patent
gb_sts:CNS06FCD	-	8.00	132.96	48.25	972	! AL396227 T7 end of clone AR0AA0
em_htgo_inv:AC043561	+	8.00	132.73	49.70	1005	! Ac043561 Giardia intestinalis
gb_pl:AY035150	+	8.00	132.51	51.14	1038	! AY035150 Arabidopsis thalia
em_htgo_inv:AC046945	-	8.00	132.20	53.22	1086	! Ac046945 Giardia intestinal
gb_pl:ZFL012562	+	8.00	131.92	55.12	1130	! AJ012562 Zamia floridiana c
gb_pl:SCSSO2A	-	8.00	131.38	59.11	1223	! X67730 S.cerevisiae SSO2 mR
gb_ro:AF156797	+	8.00	126.60	109.13	2449	! AF156797 Mus musculus ATP s
gb_ov:AY034614	+	8.00	126.32	113.10	2550	! AY034614 Danio rerio Sec61
gb_pr:AK027492	+	8.00	126.01	117.67	2667	! AK027492 Homo sapiens cDNA

seq_name: gb_pat:AX081233

seq_documentation_block:
LOCUS AX081233 336 bp DNA linear PAT 27-FEB-2001
DEFINITION Sequence 1 from Patent WO0109337.
ACCESSION AX081233
VERSION AX081233.1 GI:13170101
KEYWORDS .
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 336)
AUTHORS Thonnard,J.S.
TITLE Moraxella catarrhalis antigens basb122 and basb124
JOURNAL Patent: WO 0109337-A 1 08-FEB-2001;
SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)
FEATURES Location/Qualifiers
source 1. .336
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 106 a 65 c 61 g 104 t
ORIGIN

alignment_scores:
Quality: 111.00 Length: 111
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AX081233 ..

Align seg 1/1 to: AX081233 from: 1 to: 336

```
      1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
        |||||||||||||||||||||||||||||||||||||||||||||||||||
      1 ATGTTAAATCATCATATTCGCCTGACTATTTCCGCTTTACTGACGGCACT 50

     17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnA 34
        |||||||||||||||||||||||||||||||||||||||||||||||||||
     51 TTTGGTAACAGGTTGTGTTTCTACTGGTAATGTTGCAATGAAAGAGCAAA 100

     34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    101 ACCAACAAACCATTGAGCAGACCATCATTAAGGGCAAGACCAATAAGCAA 150

     51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVa 67
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    151 GAGATTTCTAGTAGATTTGGTTCTGCTGATAGCATCTCTTTTATGATAGT 200

     67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnG 84
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    201 GGTAATAAAATTTGGACATACCGCCATACTCGCTCCAAACCGATGGCAAG 250

     84 luIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
        |||||||||||||||||||||||||||||||||||||||||||||||||||
    251 AAATTTTATCCCTTATAATTTCTTTTCTTTGGGTGAAACCATACAGACCA 300

    101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
        |||||||||||||||||||||||||||||||||
    301 AAGAACTTGTCATTTTATTTGACAGCAAAGGCA 333
```

seq_name: gb_pat:AX067463

```
seq_documentation_block:
LOCUS       AX067463               94750 bp    DNA     linear   PAT 24-JAN-2001
DEFINITION  Sequence 38 from Patent WO0078968.
ACCESSION   AX067463
VERSION     AX067463.1  GI:12545083
KEYWORDS    .
SOURCE      Moraxella catarrhalis.
  ORGANISM  Moraxella catarrhalis
            Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
            Moraxella.
REFERENCE   1  (bases 1 to 94750)
  AUTHORS   Lagace,R.E., Patterson,C. and Berg,K.L.
  TITLE     Nucleotide sequences of moraxella catarrhalis genome
  JOURNAL   Patent: WO 0078968-A 38 28-DEC-2000;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .94750
                     /organism="Moraxella catarrhalis"
                     /db_xref="taxon:480"
BASE COUNT    27128 a  18811 c  21444 g  27367 t
ORIGIN


alignment_scores:
            Quality:  111.00              Length:    111
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AX067463   ..

 Align seg 1/1  to: AX067463  from: 1  to: 94750

       1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75267 ATGTTAAATCATCATATTCGCCTGACTATTTCCGCTTTACTGACGGCACT 75316

      17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75317 TTTGGTAACAGGTTGTGTTTCTACTGGTAATGTTGCAATGAAAGAGCAAA 75366

      34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75367 ACCAACAAACCATTGAGCAGACCATCATTAAGGGCAAGACCAATAAGCAA 75416

      51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVa 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75417 GAGATTTCTAGTAGATTTGGTTCTGCTGATAGCATCTCTTTTATGATAGT 75466

      67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75467 GGTAATAAAATTTGGACATACCGCCATACTCGCTCCAAACCGATGGCAAG 75516

      84 luIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75517 AAATTTTATCCCTTATAATTTCTTTTCTTTGGGTGAAACCATACAGACCA 75566

     101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
         |||||||||||||||||||||||||||||||
   75567 AAGAACTTGTCATTTTATTTGACAGCAAAGGCA 75599

seq_name: gb_htg:AC099920

seq_documentation_block:
LOCUS       AC099920               48309 bp    DNA     linear   HTG 22-NOV-2001
DEFINITION  Mus musculus clone RP23-13M14, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC099920
VERSION     AC099920.1  GI:17047286
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 48309)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-13M14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 48309)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L13528
                Center clone name: 13_M_14
            --------------.
            * NOTE: This record contains 61 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      678: contig of 678 bp in length
            *      679 778: gap of      100 bp
            *      779     1501: contig of 723 bp in length
            *     1502 1601: gap of      100 bp
            *     1602     2292: contig of 691 bp in length
            *     2293 2392: gap of      100 bp
            *     2393     3082: contig of 690 bp in length
            *     3083 3182: gap of      100 bp
            *     3183     3831: contig of 649 bp in length
            *     3832 3931: gap of      100 bp
            *     3932     4480: contig of 549 bp in length
            *     4481 4580: gap of      100 bp
            *     4581     5283: contig of 703 bp in length
            *     5284 5383: gap of      100 bp
            *     5384     6063: contig of 680 bp in length
            *     6064 6163: gap of      100 bp
            *     6164     6896: contig of 733 bp in length
            *     6897 6996: gap of      100 bp
            *     6997     7641: contig of 645 bp in length
            *     7642 7741: gap of      100 bp
```

```
*     7742     8458: contig of 717 bp in length
*     8459 8558: gap of      100 bp
*     8559     9290: contig of 732 bp in length
*     9291 9390: gap of      100 bp
*     9391    10112: contig of 722 bp in length
*    10113 10212: gap of      100 bp
*    10213    10927: contig of 715 bp in length
*    10928 11027: gap of      100 bp
*    11028    11736: contig of 709 bp in length
*    11737 11836: gap of      100 bp
*    11837    12561: contig of 725 bp in length
*    12562 12661: gap of      100 bp
*    12662    13345: contig of 684 bp in length
*    13346 13445: gap of      100 bp
*    13446    14128: contig of 683 bp in length
*    14129 14228: gap of      100 bp
*    14229    14901: contig of 673 bp in length
*    14902 15001: gap of      100 bp
*    15002    15681: contig of 680 bp in length
*    15682 15781: gap of      100 bp
*    15782    16458: contig of 677 bp in length
*    16459 16558: gap of      100 bp
*    16559    17233: contig of 675 bp in length
*    17234 17333: gap of      100 bp
*    17334    18010: contig of 677 bp in length
*    18011 18110: gap of      100 bp
*    18111    18832: contig of 722 bp in length
*    18833 18932: gap of      100 bp
*    18933    19663: contig of 731 bp in length
*    19664 19763: gap of      100 bp
*    19764    20436: contig of 673 bp in length
*    20437 20536: gap of      100 bp
*    20537    21242: contig of 706 bp in length
*    21243 21342: gap of      100 bp
*    21343    22045: contig of 703 bp in length
*    22046 22145: gap of      100 bp
*    22146    22842: contig of 697 bp in length
*    22843 22942: gap of      100 bp
*    22943    23652: contig of 710 bp in length
*    23653 23752: gap of      100 bp
*    23753    24481: contig of 729 bp in length
*    24482 24581: gap of      100 bp
*    24582    25312: contig of 731 bp in length
*    25313 25412: gap of      100 bp
*    25413    26099: contig of 687 bp in length
*    26100 26199: gap of      100 bp
*    26200    26890: contig of 691 bp in length
*    26891 26990: gap of      100 bp
*    26991    27663: contig of 673 bp in length
*    27664 27763: gap of      100 bp
*    27764    28454: contig of 691 bp in length
*    28455 28554: gap of      100 bp
*    28555    29267: contig of 713 bp in length
*    29268 29367: gap of      100 bp
*    29368    30081: contig of 714 bp in length
*    30082 30181: gap of      100 bp
*    30182    30867: contig of 686 bp in length
*    30868 30967: gap of      100 bp
*    30968    31697: contig of 730 bp in length
*    31698 31797: gap of      100 bp
*    31798    32483: contig of 686 bp in length
*    32484 32583: gap of      100 bp
*    32584    33245: contig of 662 bp in length
*    33246 33345: gap of      100 bp
*    33346    34026: contig of 681 bp in length
*    34027 34126: gap of      100 bp
*    34127    34845: contig of 719 bp in length
*    34846 34945: gap of      100 bp
*    34946    35647: contig of 702 bp in length
*    35648 35747: gap of      100 bp
*    35748    36423: contig of 676 bp in length
*    36424 36523: gap of      100 bp
*    36524    37241: contig of 718 bp in length
*    37242 37341: gap of      100 bp
*    37342    38019: contig of 678 bp in length
*    38020 38119: gap of      100 bp
*    38120    38780: contig of 661 bp in length
*    38781 38880: gap of      100 bp
*    38881    39596: contig of 716 bp in length
*    39597 39696: gap of      100 bp
*    39697    40393: contig of 697 bp in length
*    40394 40493: gap of      100 bp
*    40494    41169: contig of 676 bp in length
*    41170 41269: gap of      100 bp
*    41270    41922: contig of 653 bp in length
*    41923 42022: gap of      100 bp
*    42023    42711: contig of 689 bp in length
*    42712 42811: gap of      100 bp
*    42812    43507: contig of 696 bp in length
*    43508 43607: gap of      100 bp
*    43608    44336: contig of 729 bp in length
*    44337 44436: gap of      100 bp
*    44437    45109: contig of 673 bp in length
*    45110 45209: gap of      100 bp
*    45210    45906: contig of 697 bp in length
*    45907 46006: gap of      100 bp
*    46007    46723: contig of 717 bp in length
*    46724 46823: gap of      100 bp
*    46824    47513: contig of 690 bp in length
*    47514 47613: gap of      100 bp
*    47614    48309: contig of 696 bp in length.
FEATURES             Location/Qualifiers
     source          1. .48309
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-13M14"
                     /clone_lib="RPCI-23 Female Mouse BAC"
BASE COUNT    13627 a    7821 c    7946 g   12849 t    6066 others
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC099920/rev   ..

 Align seg 1/1  to reverse of: AC099920  from: 1  to: 48309

      84 GluIleLeuSerLeuIleIleSerPhe 92
         |||||||||||||||||||||||||||||
   40205 GAGATCCTGTCTTTGATAATTTCTTTT 40179

seq_name: gb_htg:AC103410

seq_documentation_block:
LOCUS       AC103410              64235 bp    DNA     linear   HTG 24-NOV-2001
DEFINITION  Mus musculus clone RP24-363H11, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC103410
VERSION     AC103410.1  GI:17063285
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 64235)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-363H11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 64235)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
```

Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (24-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L19943
Center clone name: 363_H_11
----------.
* NOTE: This record contains 81 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 675: contig of 675 bp in length
* 676 775: gap of 100 bp
* 776 1459: contig of 684 bp in length
* 1460 1559: gap of 100 bp
* 1560 2263: contig of 704 bp in length
* 2264 2363: gap of 100 bp
* 2364 3059: contig of 696 bp in length
* 3060 3159: gap of 100 bp
* 3160 3796: contig of 637 bp in length
* 3797 3896: gap of 100 bp
* 3897 4596: contig of 700 bp in length
* 4597 4696: gap of 100 bp
* 4697 5396: contig of 700 bp in length
* 5397 5496: gap of 100 bp
* 5497 6193: contig of 697 bp in length
* 6194 6293: gap of 100 bp
* 6294 6981: contig of 688 bp in length
* 6982 7081: gap of 100 bp
* 7082 7784: contig of 703 bp in length
* 7785 7884: gap of 100 bp
* 7885 8599: contig of 715 bp in length
* 8600 8699: gap of 100 bp
* 8700 9393: contig of 694 bp in length
* 9394 9493: gap of 100 bp
* 9494 10205: contig of 712 bp in length
* 10206 10305: gap of 100 bp
* 10306 11012: contig of 707 bp in length
* 11013 11112: gap of 100 bp
* 11113 11807: contig of 695 bp in length
* 11808 11907: gap of 100 bp
* 11908 12604: contig of 697 bp in length
* 12605 12704: gap of 100 bp
* 12705 13390: contig of 686 bp in length
* 13391 13490: gap of 100 bp
* 13491 14158: contig of 668 bp in length
* 14159 14258: gap of 100 bp
* 14259 14953: contig of 695 bp in length
* 14954 15053: gap of 100 bp
* 15054 15749: contig of 696 bp in length
* 15750 15849: gap of 100 bp
* 15850 16530: contig of 681 bp in length
* 16531 16630: gap of 100 bp
* 16631 17333: contig of 703 bp in length
* 17334 17433: gap of 100 bp
* 17434 18135: contig of 702 bp in length
* 18136 18235: gap of 100 bp
* 18236 18908: contig of 673 bp in length
* 18909 19008: gap of 100 bp
* 19009 19717: contig of 709 bp in length
* 19718 19817: gap of 100 bp
* 19818 20503: contig of 686 bp in length
* 20504 20603: gap of 100 bp
* 20604 21290: contig of 687 bp in length
* 21291 21390: gap of 100 bp
* 21391 22078: contig of 688 bp in length
* 22079 22178: gap of 100 bp
* 22179 22888: contig of 710 bp in length
* 22889 22988: gap of 100 bp
* 22989 23704: contig of 716 bp in length
* 23705 23804: gap of 100 bp
* 23805 24513: contig of 709 bp in length
* 24514 24613: gap of 100 bp
* 24614 25308: contig of 695 bp in length
* 25309 25408: gap of 100 bp
* 25409 26117: contig of 709 bp in length
* 26118 26217: gap of 100 bp
* 26218 26908: contig of 691 bp in length
* 26909 27008: gap of 100 bp
* 27009 27695: contig of 687 bp in length
* 27696 27795: gap of 100 bp
* 27796 28490: contig of 695 bp in length
* 28491 28590: gap of 100 bp
* 28591 29286: contig of 696 bp in length
* 29287 29386: gap of 100 bp
* 29387 30092: contig of 706 bp in length
* 30093 30192: gap of 100 bp
* 30193 30896: contig of 704 bp in length
* 30897 30996: gap of 100 bp
* 30997 31697: contig of 701 bp in length
* 31698 31797: gap of 100 bp
* 31798 32513: contig of 716 bp in length
* 32514 32613: gap of 100 bp
* 32614 33283: contig of 670 bp in length
* 33284 33383: gap of 100 bp
* 33384 34068: contig of 685 bp in length
* 34069 34168: gap of 100 bp
* 34169 34809: contig of 641 bp in length
* 34810 34909: gap of 100 bp
* 34910 35591: contig of 682 bp in length
* 35592 35691: gap of 100 bp
* 35692 36381: contig of 690 bp in length
* 36382 36481: gap of 100 bp
* 36482 37178: contig of 697 bp in length
* 37179 37278: gap of 100 bp
* 37279 37973: contig of 695 bp in length
* 37974 38073: gap of 100 bp
* 38074 38759: contig of 686 bp in length
* 38760 38859: gap of 100 bp
* 38860 39558: contig of 699 bp in length
* 39559 39658: gap of 100 bp
* 39659 40336: contig of 678 bp in length
* 40337 40436: gap of 100 bp

```
            *    40437    41136: contig of 700 bp in length
            *    41137 41236: gap of      100 bp
            *    41237    41882: contig of 646 bp in length
            *    41883 41982: gap of      100 bp
            *    41983    42683: contig of 701 bp in length
            *    42684 42783: gap of      100 bp
            *    42784    43485: contig of 702 bp in length
            *    43486 43585: gap of      100 bp
            *    43586    44275: contig of 690 bp in length
            *    44276 44375: gap of      100 bp
            *    44376    45066: contig of 691 bp in length
            *    45067 45166: gap of      100 bp
            *    45167    45847: contig of 681 bp in length
            *    45848 45947: gap of      100 bp
            *    45948    46652: contig of 705 bp in length
            *    46653 46752: gap of      100 bp
            *    46753    47468: contig of 716 bp in length
            *    47469 47568: gap of      100 bp
            *    47569    48268: contig of 700 bp in length
            *    48269 48368: gap of      100 bp
            *    48369    49066: contig of 698 bp in length
            *    49067 49166: gap of      100 bp
            *    49167    49845: contig of 679 bp in length
            *    49846 49945: gap of      100 bp
            *    49946    50630: contig of 685 bp in length
            *    50631 50730: gap of      100 bp
            *    50731    51442: contig of 712 bp in length
            *    51443 51542: gap of      100 bp
            *    51543    52232: contig of 690 bp in length
            *    52233 52332: gap of      100 bp
            *    52333    53031: contig of 699 bp in length
            *    53032 53131: gap of      100 bp
            *    53132    53826: contig of 695 bp in length
            *    53827 53926: gap of      100 bp
            *    53927    54625: contig of 699 bp in length
            *    54626 54725: gap of      100 bp

alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC103410   ..

 Align seg 1/1  to: AC103410  from: 1  to: 64235

       8 LeuThrIleSerAlaLeuLeuThrAla 16
         ||||||||||||||||||||||||||||
   49572 TTGACCATTTCGGCGTTGTTAACGGCT 49598

seq_name: gb_pr:AC064841

seq_documentation_block:
LOCUS       AC064841               69829 bp    DNA     linear   PRI 01-FEB-2002
DEFINITION  Homo sapiens chromosome 2 clone CTD-2015A10, complete sequence.
ACCESSION   AC064841
VERSION     AC064841.3  GI:18464243
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 69829)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 69829)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 69829)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Feb 1, 2002 this sequence version replaced gi:18151025.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: H_MS2015A10
            ------------------------------------------------------
FEATURES             Location/Qualifiers
     source          1. .69829
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="CTD-2015A10"
BASE COUNT    20182 a  12890 c  12669 g  24088 t
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC064841   ..

 Align seg 1/1  to: AC064841  from: 1  to: 69829

      86 LeuSerLeuIleIleSerPheLeuTrp 94
         |||||||||||||||||||||||||||
    5553 CTTTCATTAATAATATCATTTCTTTGG 5579

seq_name: gb_htg:AL670662

seq_documentation_block:
LOCUS       AL670662              113776 bp    DNA     linear   HTG 01-FEB-2002
DEFINITION  Homo sapiens chromosome 6 clone XXbac-93B10, *** SEQUENCING IN
            PROGRESS ***, 16 unordered pieces.
ACCESSION   AL670662
VERSION     AL670662.3  GI:18491449
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 4, 2002 this sequence version replaced gi:18477451.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bPG93B10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 107323 bases at least Q40
            Consensus quality: 109309 bases at least Q30
```

```
            Consensus quality: 110694 bases at least Q20
            Insert size: 112276; sum-of-contigs
            Insert size: 118454; 4.0% error; agarose-fp
            Quality coverage: 4.94x in Q20 bases; sum-of-contigs Quality
            coverage: 4.73x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2288: contig of 2288 bp in length
            *    2289 2388: gap of      100 bp
            *    2389     5333: contig of 2945 bp in length
            *    5334 5433: gap of      100 bp
            *    5434    11933: contig of 6500 bp in length
            *   11934 12033: gap of      100 bp
            *   12034    20810: contig of 8777 bp in length
            *   20811 20910: gap of      100 bp
            *   20911    29466: contig of 8556 bp in length
            *   29467 29566: gap of      100 bp
            *   29567    32572: contig of 3006 bp in length
            *   32573 32672: gap of      100 bp
            *   32673    36440: contig of 3768 bp in length
            *   36441 36540: gap of      100 bp
            *   36541    38877: contig of 2337 bp in length
            *   38878 38977: gap of      100 bp
            *   38978    43509: contig of 4532 bp in length
            *   43510 43609: gap of      100 bp
            *   43610    57681: contig of 14072 bp in length
            *   57682 57781: gap of      100 bp
            *   57782    72280: contig of 14499 bp in length
            *   72281 72380: gap of      100 bp
            *   72381    86911: contig of 14531 bp in length
            *   86912 87011: gap of      100 bp
            *   87012    94994: contig of 7983 bp in length
            *   94995 95094: gap of      100 bp
            *   95095    98312: contig of 3218 bp in length
            *   98313 98412: gap of      100 bp
            *   98413   106958: contig of 8546 bp in length
            *  106959 107058: gap of      100 bp
            *  107059   113776: contig of 6718 bp in length.
FEATURES             Location/Qualifiers
     source          1. .113776
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="XXbac-93B10"
                     /clone_lib="CHORI-501"
     misc_feature    1. .2288
                     /note="assembly_fragment:00049
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    2389. .5333
                     /note="assembly_fragment:00859
                     fragment_chain:1"
     misc_feature    5434. .11933
                     /note="assembly_fragment:01130
                     fragment_chain:1"
     misc_feature    12034. .20810
                     /note="assembly_fragment:00562
                     fragment_chain:2"
     misc_feature    20911. .29466
                     /note="assembly_fragment:00633
                     fragment_chain:2"
     misc_feature    29567. .32572
                     /note="assembly_fragment:00674
                     fragment_chain:2"
     misc_feature    32673. .36440
                     /note="assembly_fragment:00655
                     fragment_chain:2"
     misc_feature    36541. .38877
                     /note="assembly_fragment:00599
                     fragment_chain:3"
     misc_feature    38978. .43509
                     /note="assembly_fragment:00313
                     fragment_chain:3"
     misc_feature    43610. .57681
                     /note="assembly_fragment:00838
                     fragment_chain:3"
     misc_feature    57782. .72280
                     /note="assembly_fragment:00179
                     fragment_chain:4"
     misc_feature    72381. .86911
                     /note="assembly_fragment:00755
                     fragment_chain:4"
     misc_feature    87012. .94994
                     /note="assembly_fragment:00446
                     fragment_chain:5"
     misc_feature    95095. .98312
                     /note="assembly_fragment:01230
                     fragment_chain:5"
     misc_feature    98413. .106958
                     /note="assembly_fragment:00654"
     misc_feature    107059. .113776
                     /note="assembly_fragment:01253
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    30735 a  25968 c  25314 g  30248 t   1511 others
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AL670662   ..

 Align seg 1/1  to: AL670662  from: 1  to: 113776

     101 LysAsnLeuSerPheTyrLeuThrAla 109
         |||||||||||||||||||||||||||
   64868 AAAAACCTCAGTTTCTACTTGACAGCA 64894

seq_name: gb_pr:AF241734

seq_documentation_block:
LOCUS       AF241734              119919 bp    DNA     linear   PRI 05-APR-2000
DEFINITION  Homo sapiens PAC M1034 chromosome X map Xp11.4, complete sequence.
ACCESSION   AF241734
VERSION     AF241734.1  GI:7417369
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 119919)
  AUTHORS   Blechschmidt,K., Nyakatura,G., Menzel,U., Baumgart,C., Dette,M.,
            Jahn,N., Strom,T.M., Hellebrand,H., Meindl,A. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-2000) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .119919
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xp11.4"
                     /clone="PAC M1034"
     repeat_region   1221. .1351
```

```
                     /note="(CATA)n"
                     /rpt_family="Simple_repeat"
exon                 1679. .1785
                     /note="GRAIL"
                     /evidence=not_experimental
exon                 complement(3238. .3486)
                     /note="GRAIL"
                     /evidence=not_experimental
exon                 complement(3247. .3329)
                     /note="XPOUND"
                     /evidence=not_experimental
exon                 3715. .3857
                     /note="MZEF"
                     /evidence=not_experimental
exon                 complement(4848. .5069)
                     /note="GRAIL"
                     /evidence=not_experimental
exon                 complement(4890. .5039)
                     /note="MZEF"
                     /evidence=not_experimental
exon                 complement(4890. .5069)
                     /note="GENSCAN"
exon                 complement(5114. .5216)
                     /note="XPOUND"
                     /evidence=not_experimental
repeat_region        complement(5119. .5249)
                     /note="(GA)n"
                     /rpt_family="Simple_repeat"
repeat_region        complement(5438. .5615)
                     /rpt_family="LTR16A"
exon                 complement(5655. .5735)
                     /note="GRAIL"
                     /evidence=not_experimental
repeat_region        5781. .5898
                     /rpt_family="MLT1I"
repeat_region        complement(5941. .6025)
                     /rpt_family="MER58A"
exon                 complement(6370. .6471)
                     /note="GRAIL"
                     /evidence=not_experimental
exon                 complement(6370. .6471)
                     /note="GENSCAN"
exon                 6373. .6474
                     /note="GRAIL"
                     /evidence=not_experimental
exon                 complement(6589. .6652)
                     /note="XPOUND"
                     /evidence=not_experimental
repeat_region        6911. .7196
                     /rpt_family="LTR16A"
repeat_region        7223. .7995
                     /rpt_family="L1M4_orf2"
repeat_region        7847. .8729
                     /rpt_family="L1MB8"
repeat_region        8730. .8799
                     /rpt_family="LTR16A"
exon                 complement(9152. .9290)
                     /note="GRAIL"
                     /evidence=not_experimental
repeat_region        complement(9577. .9684)
                     /rpt_family="L2b"
repeat_region        9901. .10097
                     /rpt_family="MER3"
repeat_region        10692. .10866
                     /rpt_family="L2a"
repeat_region        10756. .10866
                     /rpt_family="L2b"
exon                 complement(11684. .11746)
                     /note="GRAIL"
                     /evidence=not_experimental
repeat_region        complement(12928. .13387)
                     /rpt_family="MER5A"
repeat_region        12973. .13271
                     /rpt_family="AluSx"
exon                 13276. .13386
                     /note="MZEF"
                     /evidence=not_experimental
repeat_region        13568. .13666
                     /note="(GGAA)n"
                     /rpt_family="Simple_repeat"
repeat_region        14381. .14523
                     /rpt_family="MER5A"
repeat_region        14533. .14565
                     /rpt_family="Charlie4"
repeat_region        15505. .15817
                     /rpt_family="AluSx"
repeat_region        complement(16868. .16963)
                     /rpt_family="MLT1J"
repeat_region        complement(17112. .17593)
                     /rpt_family="L2"
exon                 17967. .18107
                     /note="GRAIL"
                     /evidence=not_experimental
exon                 complement(17967. .18099)
                     /note="GRAIL"
                     /evidence=not_experimental
repeat_region        18005. .18291
                     /rpt_family="L1MB6"
repeat_region        18005. .18256
                     /rpt_family="L1ME2"
exon                 18163. .18253
                     /note="GRAIL"
                     /evidence=not_experimental
repeat_region        18309. .18559
                     /rpt_family="L1ME3"
exon                 18806. .18897
                     /note="GRAIL"
                     /evidence=not_experimental
repeat_region        complement(18918. .19221)
                     /rpt_family="AluSx"
repeat_region        20125. .20201
                     /rpt_family="LTR33"
repeat_region        20444. .22837
                     /rpt_family="LTR33"
repeat_region        complement(20514. .20871)
                     /rpt_family="THE1B"
repeat_region        complement(20872. .22434)
                     /rpt_family="THE1-INTERNAL"
exon                 21254. .21350
                     /note="GENSCAN"
exon                 complement(22026. .22156)
                     /note="MZEF"
                     /evidence=not_experimental
exon                 complement(22041. .22247)
                     /note="GRAIL"
                     /evidence=not_experimental
repeat_region        complement(22435. .22792)
                     /rpt_family="THE1B"
repeat_region        complement(23324. .23424)
                     /rpt_family="MIR"
exon                 23464. .23516
                     /note="GRAIL"
                     /evidence=not_experimental
exon                 23464. .23522
                     /note="GENSCAN"
repeat_region        complement(23649. .23713)
                     /rpt_family="MSTC"
repeat_region        complement(23735. .25349)
                     /rpt_family="MST-INTERNAL"
exon                 24441. .24528
                     /note="GRAIL"
                     /evidence=not_experimental
repeat_region        complement(25350. .25742)
                     /rpt_family="MLT1A1"
exon                 25430. .25529
                     /note="MZEF"
```

```
                     /evidence=not_experimental
     repeat_region   complement(26059. .27233)
                     /rpt_family="L1MA6"
     repeat_region   complement(26414. .26707)
                     /rpt_family="AluJb"
     repeat_region   27233. .27595
                     /rpt_family="LTR16C"
     repeat_region   28089. .28802
                     /rpt_family="L1MB3"
     repeat_region   complement(28911. .29028)
                     /rpt_family="AluJo"
     exon            28914. .29057
                     /note="MZEF"
                     /evidence=not_experimental
     repeat_region   complement(29133. .29404)
                     /rpt_family="LTR16A"
     repeat_region   29635. .29796
                     /rpt_family="MIR"
     repeat_region   complement(29991. .30286)
                     /rpt_family="AluSx"
     repeat_region   complement(30538. .30579)
                     /note="(CAAAAA)n"
                     /rpt_family="Simple_repeat"
     exon            complement(30917. .31027)
                     /note="GRAIL"
                     /evidence=not_experimental
     repeat_region   31215. .31363
                     /rpt_family="MER5A"
     exon            complement(31310. .31450)
                     /note="GRAIL"
                     /evidence=not_experimental

alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AF241734   ..

 Align seg 1/1  to: AF241734  from: 1  to: 119919

     101 LysAsnLeuSerPheTyrLeuThrAla 109
         ||||||||||||||||||||||||||||
   39908 AAAAACCTCAGTTTCTACTTGACAGCA 39934

seq_name: gb_htg:AC103115

seq_documentation_block:
LOCUS       AC103115              161924 bp    DNA     linear   HTG 21-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-188D19, *** SEQUENCING IN PROGRESS
            ***, 64 unordered pieces.
ACCESSION   AC103115
VERSION     AC103115.2  GI:17974602
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 161924)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Ogùh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161924)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 21, 2001 this sequence version replaced gi:17062758.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GJFT
                Center clone name: CH230-188D19
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 132307 bases at least Q40
                Consensus quality: 140655 bases at least Q30
                Consensus quality: 147116 bases at least Q20
                Estimated insert size: 131552; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 64 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7625: contig of 7625 bp in length
            *     7626     7725: gap of unknown length
            *     7726    14755: contig of 7030 bp in length
            *    14756    14855: gap of unknown length
            *    14856    23333: contig of 8478 bp in length
            *    23334    23433: gap of unknown length
            *    23434    27343: contig of 3910 bp in length
            *    27344    27443: gap of unknown length
            *    27444    30836: contig of 3393 bp in length
```

```
*    30837    30936: gap of unknown length
*    30937    35060: contig of 4124 bp in length
*    35061    35160: gap of unknown length
*    35161    40254: contig of 5094 bp in length
*    40255    40354: gap of unknown length
*    40355    44541: contig of 4187 bp in length
*    44542    44641: gap of unknown length
*    44642    49051: contig of 4410 bp in length
*    49052    49151: gap of unknown length
*    49152    52529: contig of 3378 bp in length
*    52530    52629: gap of unknown length
*    52630    56269: contig of 3640 bp in length
*    56270    56369: gap of unknown length
*    56370    59272: contig of 2903 bp in length
*    59273    59372: gap of unknown length
*    59373    63381: contig of 4009 bp in length
*    63382    63481: gap of unknown length
*    63482    66550: contig of 3069 bp in length
*    66551    66650: gap of unknown length
*    66651    68857: contig of 2207 bp in length
*    68858    68957: gap of unknown length
*    68958    73300: contig of 4343 bp in length
*    73301    73400: gap of unknown length
*    73401    75764: contig of 2364 bp in length
*    75765    75864: gap of unknown length
*    75865    78928: contig of 3064 bp in length
*    78929    79028: gap of unknown length
*    79029    81347: contig of 2319 bp in length
*    81348    81447: gap of unknown length
*    81448    83030: contig of 1583 bp in length
*    83031    83130: gap of unknown length
*    83131    85382: contig of 2252 bp in length
*    85383    85482: gap of unknown length
*    85483    87766: contig of 2284 bp in length
*    87767    87866: gap of unknown length
*    87867    89888: contig of 2022 bp in length
*    89889    89988: gap of unknown length
*    89989    91866: contig of 1878 bp in length
*    91867    91966: gap of unknown length
*    91967    94685: contig of 2719 bp in length
*    94686    94785: gap of unknown length
*    94786    96401: contig of 1616 bp in length
*    96402    96501: gap of unknown length
*    96502    98769: contig of 2268 bp in length
*    98770    98869: gap of unknown length
*    98870   100269: contig of 1400 bp in length
*   100270   100369: gap of unknown length
*   100370   101921: contig of 1552 bp in length
*   101922   102021: gap of unknown length
*   102022   103110: contig of 1089 bp in length
*   103111   103210: gap of unknown length
*   103211   104666: contig of 1456 bp in length
*   104667   104766: gap of unknown length
*   104767   107055: contig of 2289 bp in length
*   107056   107155: gap of unknown length
*   107156   108944: contig of 1789 bp in length
*   108945   109044: gap of unknown length
*   109045   110910: contig of 1866 bp in length
*   110911   111010: gap of unknown length
*   111011   112762: contig of 1752 bp in length
*   112763   112862: gap of unknown length
*   112863   114085: contig of 1223 bp in length
*   114086   114185: gap of unknown length
*   114186   115740: contig of 1555 bp in length
*   115741   115840: gap of unknown length
*   115841   117251: contig of 1411 bp in length
*   117252   117351: gap of unknown length
*   117352   119708: contig of 2357 bp in length
*   119709   119808: gap of unknown length
*   119809   121548: contig of 1740 bp in length
*   121549   121648: gap of unknown length
*   121649   122930: contig of 1282 bp in length
*   122931   123030: gap of unknown length
*   123031   125145: contig of 2115 bp in length
*   125146   125245: gap of unknown length
*   125246   127372: contig of 2127 bp in length
*   127373   127472: gap of unknown length
*   127473   129619: contig of 2147 bp in length
*   129620   129719: gap of unknown length
*   129720   131383: contig of 1664 bp in length
*   131384   131483: gap of unknown length
*   131484   132714: contig of 1231 bp in length
*   132715   132814: gap of unknown length
*   132815   134391: contig of 1577 bp in length
*   134392   134491: gap of unknown length
*   134492   135742: contig of 1251 bp in length
*   135743   135842: gap of unknown length
*   135843   137997: contig of 2155 bp in length
*   137998   138097: gap of unknown length
*   138098   140078: contig of 1981 bp in length
*   140079   140178: gap of unknown length
*   140179   141589: contig of 1411 bp in length
*   141590   141689: gap of unknown length
*   141690   143025: contig of 1336 bp in length
*   143026   143125: gap of unknown length
*   143126   144267: contig of 1142 bp in length
*   144268   144367: gap of unknown length
*   144368   145839: contig of 1472 bp in length
*   145840   145939: gap of unknown length
*   145940   147839: contig of 1900 bp in length
*   147840   147939: gap of unknown length
*   147940   149476: contig of 1537 bp in length
*   149477   149576: gap of unknown length

alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC103115/rev   ..

 Align seg 1/1  to reverse of: AC103115  from: 1  to: 161924

      85 IleLeuSerLeuIleIleSerPheLeu 93
         |||||||||||||||||||||||||||
   20164 ATTTTAAGTCTTATTATATCGTTTCTT 20138

seq_name: em_htg_hum:AC019286

seq_documentation_block:
ID   AC019286   standard; DNA; HTG; 163134 BP.
XX
AC   AC019286;
XX
SV   AC019286.4
XX
DT   02-JAN-2000 (Rel. 62, Created)
DT   05-APR-2000 (Rel. 63, Last updated, Version 4)
XX
DE   Homo sapiens clone RP11-21D3, WORKING DRAFT SEQUENCE, 16 unordered pieces.
XX
KW   HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-163134
RA   Birren B., Linton L., Nusbaum C., Lander E.;
RT   "Homo sapiens, clone RP11-21D3";
RL   Unpublished.
XX
RN   [2]
RP   1-163134
```

```
RA   Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
RA   Anderson S., Baldwin J., Barna N., Beckerly R., Beda F., Boguslavkiy L.,
RA   Boukhgalter B., Brown A., Burkett G., Castle A., Choepel Y., Colangelo M.,
RA   Collins S., Collymore A., Cooke P., DeArellano K., Dewar K., Domino M.,
RA   Doyle M., Fenestor J., Ferreira P., FitzHugh W., Forrest C., Gage D.,
RA   Galagan J., Gardyna S., Grant G., Hagos B., Heaford A., Horton L.,
RA   Howland J.C., Johnson R., Jones C., Kann L., Karatas A., Klein J.,
RA   Landers T., Lehoczky J., Levine R., Lieu C., Liu G., Locke K.,
RA   Macdonald P., Marquis N., McEwan P., McGurk A., McKernan K., McPheeters R.,
RA   Meldrim J., Meneus L., Morrow J., Naylor J., Norman C.H., O'Connor T.,
RA   O'Donnell P., Olivar T.M., Peterson K., Pierre N., Pisani C., Pollara V.,
RA   Raymond C., Riley R., Rothman D., Roy A., Santos R., Severy P., Spencer B.,
RA   Stange-Thomann N., Stojanovic N., Subramanian A., Talamas J., Tesfaye S.,
RA   Theodore J., Tirrell A., Vassiliev H., Viel R., Vo A., Wu X., Wyman D.,
RA   Ye W.J., Zimmer A., Zody M.;
RT   ;
RL   Submitted (31-DEC-1999) to the EMBL/GenBank/DDBJ databases.
RL   Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
RL   Cambridge, MA 02141, USA
XX
CC   On Apr 1, 2000 this sequence version replaced gi:6721342.
CC   All repeats were identified using RepeatMasker:
CC   Smit, A.F.A. & Green, P. (1996-1997)
CC   http://ftp.genome.washington.edu/RM/RepeatMasker.html
CC   -------------- Genome Center
CC   Center: Whitehead Institute/ MIT Center for Genome Research
CC   Center code: WIBR
CC   Web site: http://www-seq.wi.mit.edu
CC   Contact: sequence_submissions@genome.wi.mit.edu
CC   -------------- Project Information
CC   Center project name: L3987
CC   Center clone name: 21_D_3
CC   -------------- Summary Statistics
CC   Sequencing vector: M13; M77815; 100% of reads
CC   Chemistry: Dye-terminator Big Dye; 100% of reads
CC   Assembly program: Phrap; version 0.960731
CC   Consensus quality: 155438 bases at least Q40
CC   Consensus quality: 159063 bases at least Q30
CC   Consensus quality: 160301 bases at least Q20
CC   Insert size: 158000; agarose-fp
CC   Insert size: 161634; sum-of-contigs
CC   Quality coverage: 4.8 in Q20 bases; agarose-fp
CC   Quality coverage: 4.7 in Q20 bases; sum-of-contigs
CC   --------------.
CC   * NOTE: This is a 'working draft' sequence. It currently
CC   * consists of 16 contigs. The true order of the pieces
CC   * is not known and their order in this sequence record is
CC   * arbitrary. Gaps between the contigs are represented as
CC   * runs of N, but the exact sizes of the gaps are unknown.
CC   * This record will be updated with the finished sequence
CC   * as soon as it is available and the accession number will
CC   * be preserved.
CC   *        1      823: contig of 823 bp in length
CC   *      824 923: gap of      100 bp
CC   *      924     2980: contig of 2057 bp in length
CC   *     2981 3080: gap of      100 bp
CC   *     3081     5049: contig of 1969 bp in length
CC   *     5050 5149: gap of      100 bp
CC   *     5150     7713: contig of 2564 bp in length
CC   *     7714 7813: gap of      100 bp
CC   *     7814    10311: contig of 2498 bp in length
CC   *    10312 10411: gap of      100 bp
CC   *    10412    15992: contig of 5581 bp in length
CC   *    15993 16092: gap of      100 bp
CC   *    16093    22474: contig of 6382 bp in length
CC   *    22475 22574: gap of      100 bp
CC   *    22575    29644: contig of 7070 bp in length
CC   *    29645 29744: gap of      100 bp
CC   *    29745    38378: contig of 8634 bp in length
CC   *    38379 38478: gap of      100 bp
CC   *    38479    46868: contig of 8390 bp in length
CC   *    46869 46968: gap of      100 bp
CC   *    46969    56724: contig of 9756 bp in length
CC   *    56725 56824: gap of      100 bp
CC   *    56825    68013: contig of 11189 bp in length
CC   *    68014 68113: gap of      100 bp
CC   *    68114    80779: contig of 12666 bp in length
CC   *    80780 80879: gap of      100 bp
CC   *    80880    94821: contig of 13942 bp in length
CC   *    94822 94921: gap of      100 bp
CC   *    94922   123346: contig of 28425 bp in length
CC   *   123347 123446: gap of      100 bp
CC   *   123447   163134: contig of 39688 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .163134
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT                   /clone="RP11-21D3"
FT                   /clone_lib="RPCI-11 Human Male BAC"
FT   misc_feature    1. .823
FT                   /note="assembly_fragment~clone_end:T7~vector_side:right"
FT   misc_feature    924. .2980
FT                   /note="assembly_fragment"
FT   misc_feature    3081. .5049
FT                   /note="assembly_fragment"
FT   misc_feature    5150. .7713
FT                   /note="assembly_fragment"
FT   misc_feature    7814. .10311
FT                   /note="assembly_fragment"
FT   misc_feature    10412. .15992
FT                   /note="assembly_fragment"
FT   misc_feature    16093. .22474
FT                   /note="assembly_fragment"
FT   misc_feature    22575. .29644
FT                   /note="assembly_fragment"
FT   misc_feature    29745. .38378
FT                   /note="assembly_fragment"
FT   misc_feature    38479. .46868
FT                   /note="assembly_fragment"
FT   misc_feature    46969. .56724
FT                   /note="assembly_fragment"
FT   misc_feature    56825. .68013
FT                   /note="assembly_fragment"
FT   misc_feature    68114. .80779
FT                   /note="assembly_fragment"
FT   misc_feature    80880. .94821
FT                   /note="assembly_fragment"
FT   misc_feature    94922. .123346
FT                   /note="assembly_fragment~clone_end:SP6~vector_side:right"
FT   misc_feature    123447. .163134
FT                   /note="assembly_fragment"
XX
SQ   Sequence 163134 BP; 43585 A; 37224 C; 37165 G; 43658 T; 1502 other;


alignment_scores:
           Quality:    9.00              Length:      9
             Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC019286   ..

 Align seg 1/1  to: AC019286  from: 1  to: 163134

     101 LysAsnLeuSerPheTyrLeuThrAla 109
         |||||||||||||||||||||||||||
  149030 AAAAACCTCAGTTTCTACTTGACAGCA 149056

seq_name: gb_pr:AC108881

seq_documentation_block:
LOCUS       AC108881              163230 bp    DNA     linear   PRI 01-FEB-2002
DEFINITION  Homo sapiens chromosome X clone RP11-21D3 map Xp11.4, complete
```

```
            sequence.
ACCESSION   AC108881 AC019286
VERSION     AC108881.1  GI:18464042
KEYWORDS    HTG; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163230)
  AUTHORS   Wen,G., Ramser,J., Baumgart,C., Meindl,A. and Platzer,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163230)
  AUTHORS   Wen,G. and Platzer,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT     On Feb 1, 2002 this sequence version replaced gi:7382151.
            On Jan 22, 2002 this submission replaced accession AC019286
            Drafting Center: Whitehead Institute/ MIT Center for Genome
            Research
            Finishing Center: Institute of Molecular Biotechnology
            -------------- Genome Center
                 Center: Insitute of Molecular Biotechnology
                 Center code: IMB
                 Web site: http://genome.imb-jena.de/
                 Contact: gscj-submit@genome.imb-jena.de
            -------------- Project Information
                 Center project name: X154
                 Center clone name: RP11-21D3
            -------------- Summary Statistics
                 Sequencing vector: M13; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990329
                 Consensus quality: 158563 bases at least Q40
                 Consensus quality: 160932 bases at least Q30
                 Consensus quality: 161821 bases at least Q20
                 Quality coverage: 5.31 x in Q20 bases; sum-of-contigs
            --------------.
             Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality 10.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .163230
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xp11.4"
                     /clone="RP11-21D3"
BASE COUNT    45825 a  38652 c  36567 g  42186 t
ORIGIN


alignment_scores:
           Quality:    9.00              Length:      9
             Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC108881   ..

 Align seg 1/1  to: AC108881  from: 1  to: 163230

     101 LysAsnLeuSerPheTyrLeuThrAla 109
         |||||||||||||||||||||||||||
  124373 AAAAACCTCAGTTTCTACTTGACAGCA 124399
```

```
seq_name: em_htg_hum:AC055844

seq_documentation_block:
ID   AC055844   standard; DNA; HTG; 164017 BP.
XX
AC   AC055844;
XX
SV   AC055844.2
XX
DT   20-APR-2000 (Rel. 63, Created)
DT   26-MAY-2000 (Rel. 63, Last updated, Version 2)
XX
DE   Homo sapiens chromosome 4 clone RP11-692K15 map 4, WORKING DRAFT SEQUENCE,
DE   20 unordered pieces.
XX
KW   HTG; HTGS_DRAFT; HTGS_PHASE1.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-164017
RA   Birren B., Linton L., Nusbaum C., Lander E.;
RT   "Homo sapiens chromosome 4, clone RP11-692K15";
RL   Unpublished.
XX
RN   [2]
RP   1-164017
RA   Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
RA   Anderson S., Baldwin J., Barna N., Bastien V., Beda F., Boguslavkiy L.,
RA   Boukhgalter B., Brown A., Burkett G., Campopiano A., Castle A., Choepel Y.,
RA   Colangelo M., Collins S., Collymore A., Cooke P., DeArellano K., Dewar K.,
RA   Diaz J.S., Dodge S., Domino M., Doyle M., Ferreira P., FitzHugh W.,
RA   Gage D., Galagan J., Gardyna S., Ginde S., Goyette M., Graham L.,
RA   Grand-Pierre N., Grant G., Hagos B., Heaford A., Horton L., Howland J.C.,
RA   Iliev I., Johnson R., Jones C., Kann L., Karatas A., Klein J., LaRocque K.,
RA   Lamazares R., Landers T., Lehoczky J., Levine R., Lieu C., Liu G.,
RA   Locke K., Macdonald P., Marquis N., McCarthy M., McEwan P., McGurk A.,
RA   McKernan K., McPheeters R., Meldrim J., Meneus L., Mihova T., Miranda C.,
RA   Mlenga V., Morrow J., Murphy T., Naylor J., Norman C.H., O'Connor T.,
RA   O'Donnell P., O'Neil D., Olivar T.M., Oliver J., Peterson K., Pierre N.,
RA   Pisani C., Pollara V., Raymond C., Riley R., Rogov P., Rothman D., Roy A.,
RA   Santos R., Schauer S., Severy P., Spencer B., Stange-Thomann N.,
RA   Stojanovic N., Subramanian A., Talamas J., Tesfaye S., Theodore J.,
RA   Tirrell A., Travers M., Trigilio J., Vassiliev H., Viel R., Vo A.,
RA   Wilson B., Wu X., Wyman D., Ye W.J., Young G., Zainoun J., Zimmer A.,
RA   Zody M.;
RT   ;
RL   Submitted (18-APR-2000) to the EMBL/GenBank/DDBJ databases.
RL   Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
RL   Cambridge, MA 02141, USA
XX
CC   On May 25, 2000 this sequence version replaced gi:7582677.
CC   All repeats were identified using RepeatMasker:
CC   Smit, A.F.A. & Green, P. (1996-1997)
CC   http://ftp.genome.washington.edu/RM/RepeatMasker.html
CC   -------------- Genome Center
CC   Center: Whitehead Institute/ MIT Center for Genome Research
CC   Center code: WIBR
CC   Web site: http://www-seq.wi.mit.edu
CC   Contact: sequence_submissions@genome.wi.mit.edu
CC   -------------- Project Information
CC   Center project name: L9704
CC   Center clone name: 692_K_15
CC   -------------- Summary Statistics
CC   Sequencing vector: M13; M77815; 100% of reads
CC   Chemistry: Dye-terminator Big Dye; 100% of reads
CC   Assembly program: Phrap; version 0.960731
CC   Consensus quality: 153509 bases at least Q40
CC   Consensus quality: 159062 bases at least Q30
CC   Consensus quality: 161045 bases at least Q20
CC   Insert size: 170000; agarose-fp
```

```
CC   Insert size: 162117; sum-of-contigs
CC   Quality coverage: 4.2 in Q20 bases; agarose-fp
CC   Quality coverage: 4.4 in Q20 bases; sum-of-contigs
CC   --------------.
CC   * NOTE: This is a 'working draft' sequence. It currently
CC   * consists of 20 contigs. The true order of the pieces
CC   * is not known and their order in this sequence record is
CC   * arbitrary. Gaps between the contigs are represented as
CC   * runs of N, but the exact sizes of the gaps are unknown.
CC   * This record will be updated with the finished sequence
CC   * as soon as it is available and the accession number will
CC   * be preserved.
CC   *        1     1095: contig of 1095 bp in length
CC   *     1096 1195: gap of      100 bp
CC   *     1196     3583: contig of 2388 bp in length
CC   *     3584 3683: gap of      100 bp
CC   *     3684     5501: contig of 1818 bp in length
CC   *     5502 5601: gap of      100 bp
CC   *     5602     7199: contig of 1598 bp in length
CC   *     7200 7299: gap of      100 bp
CC   *     7300     8463: contig of 1164 bp in length
CC   *     8464 8563: gap of      100 bp
CC   *     8564    12196: contig of 3633 bp in length
CC   *    12197 12296: gap of      100 bp
CC   *    12297    14633: contig of 2337 bp in length
CC   *    14634 14733: gap of      100 bp
CC   *    14734    17931: contig of 3198 bp in length
CC   *    17932 18031: gap of      100 bp
CC   *    18032    21385: contig of 3354 bp in length
CC   *    21386 21485: gap of      100 bp
CC   *    21486    25912: contig of 4427 bp in length
CC   *    25913 26012: gap of      100 bp
CC   *    26013    33022: contig of 7010 bp in length
CC   *    33023 33122: gap of      100 bp
CC   *    33123    40700: contig of 7578 bp in length
CC   *    40701 40800: gap of      100 bp
CC   *    40801    50958: contig of 10158 bp in length
CC   *    50959 51058: gap of      100 bp
CC   *    51059    58902: contig of 7844 bp in length
CC   *    58903 59002: gap of      100 bp
CC   *    59003    71011: contig of 12009 bp in length
CC   *    71012 71111: gap of      100 bp
CC   *    71112    81941: contig of 10830 bp in length
CC   *    81942 82041: gap of      100 bp
CC   *    82042    92606: contig of 10565 bp in length
CC   *    92607 92706: gap of      100 bp
CC   *    92707   105045: contig of 12339 bp in length
CC   *   105046 105145: gap of      100 bp
CC   *   105146   126729: contig of 21584 bp in length
CC   *   126730 126829: gap of      100 bp
CC   *   126830   164017: contig of 37188 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .164017
FT                   /chromosome="4"
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
FT                   /map="4"
FT                   /clone="RP11-692K15"
FT                   /clone_lib="RPCI-11 Human Male BAC"
FT   misc_feature    1. .1095
FT                   /note="assembly_fragment"
FT   misc_feature    1196. .3583
FT                   /note="assembly_fragment"
FT   misc_feature    3684. .5501
FT                   /note="assembly_fragment"
FT   misc_feature    5602. .7199
FT                   /note="assembly_fragment"
FT   misc_feature    7300. .8463
FT                   /note="assembly_fragment clone_end:SP6 vector_side:right"
FT   misc_feature    8564. .12196
FT                   /note="assembly_fragment"
FT   misc_feature    12297. .14633
FT                   /note="assembly_fragment"
FT   misc_feature    14734. .17931
FT                   /note="assembly_fragment"
FT   misc_feature    18032. .21385
FT                   /note="assembly_fragment"
FT   misc_feature    21486. .25912
FT                   /note="assembly_fragment"
FT   misc_feature    26013. .33022
FT                   /note="assembly_fragment"
FT   misc_feature    33123. .40700
FT                   /note="assembly_fragment"
FT   misc_feature    40801. .50958
FT                   /note="assembly_fragment"
FT   misc_feature    51059. .58902
FT                   /note="assembly_fragment"
FT   misc_feature    59003. .71011
FT                   /note="assembly_fragment"
FT   misc_feature    71112. .81941
FT                   /note="assembly_fragment"
FT   misc_feature    82042. .92606
FT                   /note="assembly_fragment"
FT   misc_feature    92707. .105045
FT                   /note="assembly_fragment"
FT   misc_feature    105146. .126729
FT                   /note="assembly_fragment"
FT   misc_feature    126830. .164017
FT                   /note="assembly_fragment clone_end:T7 vector_side:right"
XX
SQ   Sequence 164017 BP; 51364 A; 29555 C; 30760 G; 50436 T; 1902 other;


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC055844    ..

 Align seg 1/1  to: AC055844  from: 1  to: 164017

      82 TrpGlnGluIleLeuSerLeuIleIle 90
         ||||||||||||||||||||||||||
   45541 TGGCAGGAGATTTTGTCTTTAATTATA 45567

seq_name: gb_htg:AC010811

seq_documentation_block:
LOCUS        AC010811              165412 bp    DNA     linear   HTG 09-FEB-2002
DEFINITION   Homo sapiens chromosome 11 clone RP11-3E17 map 11, *** SEQUENCING
             IN PROGRESS ***, 5 unordered pieces.
ACCESSION    AC010811
VERSION      AC010811.10  GI:18642864
KEYWORDS     HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 165412)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens chromosome 11, clone RP11-3E17
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 165412)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
             Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
             Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
             Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
             Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
             Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
             Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
             Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
             McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
```

```
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 9, 2002 this sequence version replaced gi:18543112.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2720
                Center clone name: 3_E_17
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    12794: contig of 12794 bp in length
            *    12795 12894: gap of      100 bp
            *    12895    68988: contig of 56094 bp in length
            *    68989 69088: gap of      100 bp
            *    69089   133991: contig of 64903 bp in length
            *   133992 134091: gap of      100 bp
            *   134092   162167: contig of 28076 bp in length
            *   162168 162267: gap of      100 bp
            *   162268   165412: contig of 3145 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165412
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-3E17"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    48029 a  30561 c  34765 g  51633 t    424 others
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC010811/rev   ..

 Align seg 1/1  to reverse of: AC010811  from: 1  to: 165412

      85 IleLeuSerLeuIleIleSerPheLeu 93
         |||||||||||||||||||||||||||
  109495 ATCCTCTCTTTAATTATCTCCTTTCTT 109469

seq_name: gb_htg:AC106051

seq_documentation_block:
LOCUS       AC106051              168577 bp    DNA     linear   HTG 11-JAN-2002
DEFINITION  Homo sapiens chromosome 4 clone RP11-692K15, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC106051 AC055844
VERSION     AC106051.1  GI:18129600
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168577)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168577)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jan 11, 2002 this sequence version replaced gi:8084355.

            ------------------- Genome Center -------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: H_NH0692K15
            Drafting center: WIBR
            ------------------------------------------------------
            -------------- Summary Statistics --------------
            Sequencing vector: M13; 52%
            Sequencing vector: plasmid; 48%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 167893 bases at least Q40
            Consensus quality: 168622 bases at least Q30
            Consensus quality: 168958 bases at least Q20
            Insert size: 170000; agarose-fp
            Insert size: 172425; sum-of-contigs
            Quality coverage: 9.25 in Q20 bases; agarose-fp
            Quality coverage: 8.49 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1210: contig of 1210 bp in length
            *     1211     1310: gap of unknown length
            *     1311    35898: contig of 34588 bp in length
            *    35899    35998: gap of unknown length
            *    35999   168577: contig of 132579 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168577
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-692K15"
     misc_feature    1. .1210
                     /note="assembly_name:Contig17"
     misc_feature    1311. .35898
                     /note="assembly_name:Contig18"
     misc_feature    35999. .168577
                     /note="assembly_name:Contig19
                     clone_end:T7
                     vector_side:right"
BASE COUNT    53348 a  31187 c  31329 g  52513 t    200 others
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
```

```
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC106051   ..

 Align seg 1/1  to: AC106051  from: 1  to: 168577

      82 TrpGlnGluIleLeuSerLeuIleIle 90
         ||||||||||||||||||||||||||
   70847 TGGCAGGAGATTTTGTCTTTAATTATA 70873

seq_name: gb_pr:AC097661

seq_documentation_block:
LOCUS       AC097661              169197 bp    DNA     linear   PRI 03-JAN-2002
DEFINITION  Homo sapiens chromosome 4 clone RP11-553E15, complete sequence.
ACCESSION   AC097661 AC023845
VERSION     AC097661.3  GI:17975439
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169197)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169197)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   3  (bases 1 to 169197)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 169197)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Dec 21, 2001 this sequence version replaced gi:17298655.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information ----------------
            Center project name: H_NH0553E15
            Drafting center: WIBR
            ----------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .169197
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-553E15"
BASE COUNT    56173 a  31106 c  29512 g  52406 t
ORIGIN


alignment_scores:
           Quality:    9.00            Length:      9
             Ratio:   1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC097661/rev   ..

 Align seg 1/1  to reverse of: AC097661  from: 1  to: 169197

      10 IleSerAlaLeuLeuThrAlaLeuLeu 18
         ||||||||||||||||||||||||||
  126364 ATTTCTGCCCTGCTTACAGCTTTACTC 126338

seq_name: gb_ro:AC074041

seq_documentation_block:
LOCUS       AC074041              170548 bp    DNA     linear   ROD 28-JUN-2001
DEFINITION  Mus musculus chromosome 2 clone RP23-258D16 strain C57BL6/J,
            complete sequence.
ACCESSION   AC074041
VERSION     AC074041.3  GI:14573692
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 170548)
  AUTHORS   Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L.,
            Granite,S., Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E.,
            Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
            Masiello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
            Pearson,R., Prasad,A., Shevchenko,Y., Snyder,B., Stantripop,S.,
            Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L.,
            Walker,M.A., Wetherby,K.D., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170548)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 170548)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jun 28, 2001 this sequence version replaced gi:11024833.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                 Center project name: ta
                 Center clone name: 258D16

            This sequence was finished as follows unless otherwise noted:
            all regions were double-stranded, sequenced with an
            alternate chemistry, or covered by high quality data
            (i.e., phred quality >= 30); an attempt was made to resolve
            all sequencing problems, such as compressions and repeats;
            all regions were covered by at least one plasmid subclone
            or more than one M13 subclone; and the assembly was confirmed
            by restriction digest.

            CLONE LENGTH: This sequence represents the entire insert of
            this clone unless otherwise noted. If there are overlapping
            clones, the overlaps are noted in the beginning and end of
            the Features section.
FEATURES             Location/Qualifiers
     source          1. .170548
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-258D16"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    1. .17547
```

```
                     /note="clone overlaps with GenBank Accession Number
                     AC074224 (nucleotides 171443-188989) clone RP23-43C3
                     (center project name uz)"
     misc_feature    170267. .170274
                     /note="low quality single stranded/single chemistry
                     region"
BASE COUNT    51882 a  37418 c  35891 g  45357 t
ORIGIN


alignment_scores:
           Quality:    9.00            Length:      9
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC074041/rev   ..

 Align seg 1/1  to reverse of: AC074041  from: 1  to: 170548

       8 LeuThrIleSerAlaLeuLeuThrAla 16
         ||||||||||||||||||||||||||
    9603 TTGACCATTTCGGCGTTGTTAACGGCT 9577

seq_name: gb_htg:AC099145

seq_documentation_block:
LOCUS       AC099145              172176 bp    DNA     linear   HTG 20-DEC-2001
DEFINITION  Rattus norvegicus clone CH230-68A17, *** SEQUENCING IN PROGRESS
            ***, 75 unordered pieces.
ACCESSION   AC099145
VERSION     AC099145.2  GI:17973022
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 172176)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172176)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 20, 2001 this sequence version replaced gi:16874639.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGYQ
                Center clone name: CH230-68A17
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 137202 bases at least Q40
                Consensus quality: 145881 bases at least Q30
                Consensus quality: 153638 bases at least Q20
                Estimated insert size: 146753; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 1.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 75 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5563: contig of 5563 bp in length
            *     5564     5663: gap of unknown length
            *     5664    11484: contig of 5821 bp in length
            *    11485    11584: gap of unknown length
            *    11585    14678: contig of 3094 bp in length
            *    14679    14778: gap of unknown length
            *    14779    18800: contig of 4022 bp in length
            *    18801    18900: gap of unknown length
            *    18901    23194: contig of 4294 bp in length
            *    23195    23294: gap of unknown length
            *    23295    26458: contig of 3164 bp in length
            *    26459    26558: gap of unknown length
            *    26559    29915: contig of 3357 bp in length
            *    29916    30015: gap of unknown length
            *    30016    34394: contig of 4379 bp in length
            *    34395    34494: gap of unknown length
            *    34495    37202: contig of 2708 bp in length
            *    37203    37302: gap of unknown length
            *    37303    41893: contig of 4591 bp in length
            *    41894    41993: gap of unknown length
            *    41994    44551: contig of 2558 bp in length
            *    44552    44651: gap of unknown length
            *    44652    46827: contig of 2176 bp in length
            *    46828    46927: gap of unknown length
            *    46928    49389: contig of 2462 bp in length
            *    49390    49489: gap of unknown length
            *    49490    52444: contig of 2955 bp in length
            *    52445    52544: gap of unknown length
            *    52545    55673: contig of 3129 bp in length
            *    55674    55773: gap of unknown length
            *    55774    58799: contig of 3026 bp in length
```

```
*    58800    58899: gap of unknown length
*    58900    60728: contig of 1829 bp in length
*    60729    60828: gap of unknown length
*    60829    64135: contig of 3307 bp in length
*    64136    64235: gap of unknown length
*    64236    66744: contig of 2509 bp in length
*    66745    66844: gap of unknown length
*    66845    70140: contig of 3296 bp in length
*    70141    70240: gap of unknown length
*    70241    72453: contig of 2213 bp in length
*    72454    72553: gap of unknown length
*    72554    75923: contig of 3370 bp in length
*    75924    76023: gap of unknown length
*    76024    77434: contig of 1411 bp in length
*    77435    77534: gap of unknown length
*    77535    80127: contig of 2593 bp in length
*    80128    80227: gap of unknown length
*    80228    82231: contig of 2004 bp in length
*    82232    82331: gap of unknown length
*    82332    83813: contig of 1482 bp in length
*    83814    83913: gap of unknown length
*    83914    86543: contig of 2630 bp in length
*    86544    86643: gap of unknown length
*    86644    87959: contig of 1316 bp in length
*    87960    88059: gap of unknown length
*    88060    91075: contig of 3016 bp in length
*    91076    91175: gap of unknown length
*    91176    92899: contig of 1724 bp in length
*    92900    92999: gap of unknown length
*    93000    95452: contig of 2453 bp in length
*    95453    95552: gap of unknown length
*    95553    96932: contig of 1380 bp in length
*    96933    97032: gap of unknown length
*    97033    99660: contig of 2628 bp in length
*    99661    99760: gap of unknown length
*    99761   101461: contig of 1701 bp in length
*   101462   101561: gap of unknown length
*   101562   103063: contig of 1502 bp in length
*   103064   103163: gap of unknown length
*   103164   104494: contig of 1331 bp in length
*   104495   104594: gap of unknown length
*   104595   107379: contig of 2785 bp in length
*   107380   107479: gap of unknown length
*   107480   110000: contig of 2521 bp in length
*   110001   110100: gap of unknown length
*   110101   112185: contig of 2085 bp in length
*   112186   112285: gap of unknown length
*   112286   114526: contig of 2241 bp in length
*   114527   114626: gap of unknown length
*   114627   116967: contig of 2341 bp in length
*   116968   117067: gap of unknown length
*   117068   118286: contig of 1219 bp in length
*   118287   118386: gap of unknown length
*   118387   119705: contig of 1319 bp in length
*   119706   119805: gap of unknown length
*   119806   120897: contig of 1092 bp in length
*   120898   120997: gap of unknown length
*   120998   123513: contig of 2516 bp in length
*   123514   123613: gap of unknown length
*   123614   125484: contig of 1871 bp in length
*   125485   125584: gap of unknown length
*   125585   128208: contig of 2624 bp in length
*   128209   128308: gap of unknown length
*   128309   129589: contig of 1281 bp in length
*   129590   129689: gap of unknown length
*   129690   131532: contig of 1843 bp in length
*   131533   131632: gap of unknown length
*   131633   133157: contig of 1525 bp in length
*   133158   133257: gap of unknown length
*   133258   135539: contig of 2282 bp in length
*   135540   135639: gap of unknown length
*   135640   136898: contig of 1259 bp in length
*   136899   136998: gap of unknown length
*   136999   138251: contig of 1253 bp in length
*   138252   138351: gap of unknown length
*   138352   139865: contig of 1514 bp in length
*   139866   139965: gap of unknown length
*   139966   141534: contig of 1569 bp in length
*   141535   141634: gap of unknown length
*   141635   143556: contig of 1922 bp in length
*   143557   143656: gap of unknown length

alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC099145   ..

 Align seg 1/1  to: AC099145  from: 1  to: 172176

      47 ThrAsnLysGlnGluIleSerSerArg 55
         |||||||||||||||||||||||||||
  126728 ACCAATAAACAAGAAATCTCAAGCAGA 126754

seq_name: gb_htg:AC013798

seq_documentation_block:
LOCUS       AC013798              175706 bp    DNA     linear   HTG 01-APR-2000
DEFINITION  Homo sapiens clone RP11-20I14, WORKING DRAFT SEQUENCE, 13 unordered
            pieces.
ACCESSION   AC013798
VERSION     AC013798.4  GI:7382144
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175706)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-20I14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175706)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 1, 2000 this sequence version replaced gi:6554527.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3908
                Center clone name: 20_I_14
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
```

```
            Assembly program: Phrap; version 0.960731
            Consensus quality: 166346 bases at least Q40
            Consensus quality: 169904 bases at least Q30
            Consensus quality: 171697 bases at least Q20
            Insert size: 183000; agarose-fp
            Insert size: 174506; sum-of-contigs
            Quality coverage: 5.2 in Q20 bases; agarose-fp
            Quality coverage: 5.5 in Q20 bases; sum-of-contigs
          ---------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 13 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1504: contig of 1504 bp in length
          *     1505  1604: gap of      100 bp
          *     1605     4186: contig of 2582 bp in length
          *     4187  4286: gap of      100 bp
          *     4287     6724: contig of 2438 bp in length
          *     6725  6824: gap of      100 bp
          *     6825    13163: contig of 6339 bp in length
          *    13164  13263: gap of      100 bp
          *    13264    18433: contig of 5170 bp in length
          *    18434  18533: gap of      100 bp
          *    18534    27908: contig of 9375 bp in length
          *    27909  28008: gap of      100 bp
          *    28009    34485: contig of 6477 bp in length
          *    34486  34585: gap of      100 bp
          *    34586    44981: contig of 10396 bp in length
          *    44982  45081: gap of      100 bp
          *    45082    56951: contig of 11870 bp in length
          *    56952  57051: gap of      100 bp
          *    57052    67256: contig of 10205 bp in length
          *    67257  67356: gap of      100 bp
          *    67357    97219: contig of 29863 bp in length
          *    97220  97319: gap of      100 bp
          *    97320   134359: contig of 37040 bp in length
          *   134360 134459: gap of      100 bp
          *   134460   175706: contig of 41247 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175706
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-20I14"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1504
                     /note="assembly_fragment"
     misc_feature    1605. .4186
                     /note="assembly_fragment"
     misc_feature    4287. .6724
                     /note="assembly_fragment"
     misc_feature    6825. .13163
                     /note="assembly_fragment"
     misc_feature    13264. .18433
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    18534. .27908
                     /note="assembly_fragment"
     misc_feature    28009. .34485
                     /note="assembly_fragment"
     misc_feature    34586. .44981
                     /note="assembly_fragment"
     misc_feature    45082. .56951
                     /note="assembly_fragment"
     misc_feature    57052. .67256
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    67357. .97219
                     /note="assembly_fragment"
     misc_feature    97320. .134359
                     /note="assembly_fragment"
     misc_feature    134460. .175706
                     /note="assembly_fragment"
BASE COUNT    46950 a  40425 c  40682 g  46442 t   1207 others
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC013798    ..

 Align seg 1/1  to: AC013798  from: 1  to: 175706

     101 LysAsnLeuSerPheTyrLeuThrAla 109
         |||||||||||||||||||||||||||
    9798 AAAAACCTCAGTTTCTACTTGACAGCA 9824

seq_name: gb_htg:AC009290

seq_documentation_block:
LOCUS        AC009290              178670 bp    DNA     linear   HTG 04-MAY-2001
DEFINITION   Homo sapiens clone RP11-45I18, WORKING DRAFT SEQUENCE, 10 unordered
             pieces.
ACCESSION    AC009290
VERSION      AC009290.3  GI:8072429
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 178670)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens, clone RP11-45I18
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 178670)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
             Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
             Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
             Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
             Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
             Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
             Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
             Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
             Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
             Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
             Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
             Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
             Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
             Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
             Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (13-AUG-1999) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On May 25, 2000 this sequence version replaced gi:7328737.
             All repeats were identified using RepeatMasker:
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             -------------- Genome Center
                 Center: Whitehead Institute/ MIT Center for Genome Research
                 Center code: WIBR
                 Web site: http://www-seq.wi.mit.edu
                 Contact: sequence_submissions@genome.wi.mit.edu
             -------------- Project Information
                 Center project name: L1165
                 Center clone name: 45_I_18
             -------------- Summary Statistics
                 Sequencing vector: M13; M77815; 93% of reads
```

```
            Sequencing vector: Plasmid; n/a; %-0.f%% of reads
            7.38747553816047Chemistry: Dye-primer-amersham; 8% of reads
            Chemistry: Dye-terminator Big Dye; 92% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 171189 bases at least Q40
            Consensus quality: 174658 bases at least Q30
            Consensus quality: 176319 bases at least Q20
            Insert size: 190000; agarose-fp
            Insert size: 177770; sum-of-contigs
            Quality coverag.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 10 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     1183: contig of 1183 bp in length
          *     1184 1283: gap of      100 bp
          *     1284     3307: contig of 2024 bp in length
          *     3308 3407: gap of      100 bp
          *     3408     7626: contig of 4219 bp in length
          *     7627 7726: gap of      100 bp
          *     7727    16657: contig of 8931 bp in length
          *    16658 16757: gap of      100 bp
          *    16758    26995: contig of 10238 bp in length
          *    26996 27095: gap of      100 bp
          *    27096    42276: contig of 15181 bp in length
          *    42277 42376: gap of      100 bp
          *    42377    60953: contig of 18577 bp in length
          *    60954 61053: gap of      100 bp
          *    61054    83707: contig of 22654 bp in length
          *    83708 83807: gap of      100 bp
          *    83808   116924: contig of 33117 bp in length
          *   116925 117024: gap of      100 bp
          *   117025   178670: contig of 61646 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178670
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-45I18"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1183
                     /note="assembly_fragment"
     misc_feature    1284. .3307
                     /note="assembly_fragment"
     misc_feature    3408. .7626
                     /note="assembly_fragment"
     misc_feature    7727. .16657
                     /note="assembly_fragment"
     misc_feature    16758. .26995
                     /note="assembly_fragment"
     misc_feature    27096. .42276
                     /note="assembly_fragment"
     misc_feature    42377. .60953
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    61054. .83707
                     /note="assembly_fragment"
     misc_feature    83808. .116924
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    117025. .178670
                     /note="assembly_fragment"
BASE COUNT    57920 a  31871 c  32262 g  55715 t    902 others
ORIGIN


alignment_scores:
          Quality:    9.00            Length:      9
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC009290/rev   ..

 Align seg 1/1  to reverse of: AC009290  from: 1  to: 178670

      86 LeuSerLeuIleIleSerPheLeuTrp 94
         ||||||||||||||||||||||||||||
   59489 CTTTCATTAATAATATCATTTCTTTGG 59463

seq_name: gb_ro:AC074224

seq_documentation_block:
LOCUS       AC074224              188989 bp    DNA     linear   ROD 22-JAN-2001
DEFINITION  Mus musculus chromosome 2 clone RP23-43C3 strain C57BL6/J, complete
            sequence.
ACCESSION   AC074224
VERSION     AC074224.3  GI:12331469
KEYWORDS    HTG.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 188989)
  AUTHORS   Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S.,
            Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E., Lee-Lin,S.-Q.,
            Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B., Masiello,C.,
            Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
            Shevchenko,Y., Snyder,B., Stantripop,S., Thomas,J.W., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188989)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 188989)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jan 22, 2001 this sequence version replaced gi:11038457.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                 Center project name: uz
                 Center clone name: 043C03

            This sequence was finished as follows unless otherwise noted:
            all regions were double-stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.

            CLONE LENGTH: This sequence represents the entire insert of
            this clone unless otherwise noted. If there are overlapping
            clones, the overlaps are noted in the beginning and end of
            the Features section.
FEATURES             Location/Qualifiers
     source          1. .188989
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
```

```
                     /chromosome="2"
                     /clone="RP23-43C3"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    75707. .75760
                     /note="pcr product sequence only"
     misc_feature    83183. .83273
                     /note="single clone coverage"
     misc_feature    103600. .103772
                     /note="single clone coverage"
BASE COUNT    53467 a  42113 c  41895 g  51514 t
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC074224/rev   ..

 Align seg 1/1  to reverse of: AC074224  from: 1  to: 188989

       8 LeuThrIleSerAlaLeuLeuThrAla 16
         ||||||||||||||||||||||||||
  181045 TTGACCATTTCGGCGTTGTTAACGGCT 181019

seq_name: gb_htg:AC023140

seq_documentation_block:
LOCUS       AC023140              194575 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-535P8, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC023140
VERSION     AC023140.4  GI:8570285
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194575)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194575)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 17, 2000 this sequence version replaced gi:7232193.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0535P08
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 183084 bases at least Q40
            Consensus quality: 186527 bases at least Q30
            Consensus quality: 188372 bases at least Q20
            Insert size: 209000; agarose-fp
            Insert size: 192075; sum-of-contigs
            Quality coverage: 3.93 in Q20 bases; agarose-fp
            Quality coverage: 4.32 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1295: contig of 1295 bp in length
            *     1296     1395: gap of unknown length
            *     1396     3355: contig of 1960 bp in length
            *     3356     3455: gap of unknown length
            *     3456     5113: contig of 1658 bp in length
            *     5114     5213: gap of unknown length
            *     5214     6910: contig of 1697 bp in length
            *     6911     7010: gap of unknown length
            *     7011     9042: contig of 2032 bp in length
            *     9043     9142: gap of unknown length
            *     9143    11645: contig of 2503 bp in length
            *    11646    11745: gap of unknown length
            *    11746    14544: contig of 2799 bp in length
            *    14545    14644: gap of unknown length
            *    14645    17322: contig of 2678 bp in length
            *    17323    17422: gap of unknown length
            *    17423    19837: contig of 2415 bp in length
            *    19838    19937: gap of unknown length
            *    19938    25695: contig of 5758 bp in length
            *    25696    25795: gap of unknown length
            *    25796    29231: contig of 3436 bp in length
            *    29232    29331: gap of unknown length
            *    29332    32892: contig of 3561 bp in length
            *    32893    32992: gap of unknown length
            *    32993    38263: contig of 5271 bp in length
            *    38264    38363: gap of unknown length
            *    38364    42404: contig of 4041 bp in length
            *    42405    42504: gap of unknown length
            *    42505    48404: contig of 5900 bp in length
            *    48405    48504: gap of unknown length
            *    48505    57290: contig of 8786 bp in length
            *    57291    57390: gap of unknown length
            *    57391    65905: contig of 8515 bp in length
            *    65906    66005: gap of unknown length
            *    66006    74728: contig of 8723 bp in length
            *    74729    74828: gap of unknown length
            *    74829    87289: contig of 12461 bp in length
            *    87290    87389: gap of unknown length
            *    87390    99467: contig of 12078 bp in length
            *    99468    99567: gap of unknown length
            *    99568   111113: contig of 11546 bp in length
            *   111114   111213: gap of unknown length
            *   111214   122406: contig of 11193 bp in length
            *   122407   122506: gap of unknown length
            *   122507   135480: contig of 12974 bp in length
            *   135481   135580: gap of unknown length
            *   135581   153618: contig of 18038 bp in length
            *   153619   153718: gap of unknown length
            *   153719   170017: contig of 16299 bp in length
            *   170018   170117: gap of unknown length
            *   170118   194575: contig of 24458 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194575
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="RP11-535P8"
     misc_feature    1. .1295
                     /note="assembly_name:Contig9"
     misc_feature    1396. .3355
                     /note="assembly_name:Contig10"
     misc_feature    3456. .5113
                     /note="assembly_name:Contig11"
     misc_feature    5214. .6910
                     /note="assembly_name:Contig12"
     misc_feature    7011. .9042
```

```
                     /note="assembly_name:Contig13"
     misc_feature    9143. .11645
                     /note="assembly_name:Contig14"
     misc_feature    11746. .14544
                     /note="assembly_name:Contig15"
     misc_feature    14645. .17322
                     /note="assembly_name:Contig16
                     clone_end:T7
                     vector_side:left"
     misc_feature    17423. .19837
                     /note="assembly_name:Contig17"
     misc_feature    19938. .25695
                     /note="assembly_name:Contig18"
     misc_feature    25796. .29231
                     /note="assembly_name:Contig19"
     misc_feature    29332. .32892
                     /note="assembly_name:Contig20"
     misc_feature    32993. .38263
                     /note="assembly_name:Contig21"
     misc_feature    38364. .42404
                     /note="assembly_name:Contig22"
     misc_feature    42505. .48404
                     /note="assembly_name:Contig23"
     misc_feature    48505. .57290
                     /note="assembly_name:Contig24"
     misc_feature    57391. .65905
                     /note="assembly_name:Contig25"
     misc_feature    66006. .74728
                     /note="assembly_name:Contig26"
     misc_feature    74829. .87289
                     /note="assembly_name:Contig27"
     misc_feature    87390. .99467
                     /note="assembly_name:Contig28"
     misc_feature    99568. .111113
                     /note="assembly_name:Contig29"
     misc_feature    111214. .122406
                     /note="assembly_name:Contig30"
     misc_feature    122507. .135480
                     /note="assembly_name:Contig31"
     misc_feature    135581. .153618
                     /note="assembly_name:Contig32"
     misc_feature    153719. .170017
                     /note="assembly_name:Contig33
                     clone_end:SP6
                     vector_side:right"
     misc_feature    170118. .194575
                     /note="assembly_name:Contig34"
BASE COUNT     58990 a  40026 c  37747 g  55237 t   2575 others
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC023140/rev   ..

 Align seg 1/1  to reverse of: AC023140  from: 1  to: 194575

      85 IleLeuSerLeuIleIleSerPheLeu 93
         ||||||||||||||||||||||||||
   73566 ATTCTGTCTCTTATCATTTCTTTCCTG 73540

seq_name: gb_htg:AC068980

seq_documentation_block:
LOCUS       AC068980              204991 bp    DNA     linear   HTG 05-OCT-2001
DEFINITION  Homo sapiens chromosome 3 clone RP11-552A14, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC068980
VERSION     AC068980.10  GI:15741365
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 204991)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 204991)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 25, 2001 this sequence version replaced gi:11128156.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HBGY
                Center clone name: RP11-552A14
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 14% of reads
                Chemistry: Dye-terminator Big Dye: 86% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 196487 bases at least Q40
                Consensus quality: 203036 bases at least Q30
                Consensus quality: 206264 bases at least Q20
                Estimated insert size: 205706; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4.6x in Q20 bases; sum-of-contigs estimation
```

```
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    52522: contig of 52522 bp in length
            *   52523    52622: gap of unknown length
            *   52623    91359: contig of 38737 bp in length
            *   91360    91459: gap of unknown length
            *   91460   125166: contig of 33707 bp in length
            *  125167   125266: gap of unknown length
            *  125267   152897: contig of 27631 bp in length
            *  152898   152997: gap of unknown length
            *  152998   173601: contig of 20604 bp in length
            *  173602   173701: gap of unknown length
            *  173702   188647: contig of 14946 bp in length
            *  188648   188747: gap of unknown length
            *  188748   202570: contig of 13823 bp in length
            *  202571   202670: gap of unknown length
            *  202671   204991: contig of 2321 bp in length.
FEATURES             Location/Qualifiers
     source          1. .204991
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-552A14"
BASE COUNT    60973 a  40617 c  40715 g  61970 t    716 others
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC068980   ..

 Align seg 1/1  to: AC068980  from: 1  to: 204991

       8 LeuThrIleSerAlaLeuLeuThrAla 16
         ||||||||||||||||||||||||||||
   65126 CTCACAATCTCTGCTCTCCTGACAGCA 65152

seq_name: gb_htg:AC024447

seq_documentation_block:
LOCUS        AC024447              204993 bp    DNA     linear   HTG 11-APR-2000
DEFINITION   Homo sapiens chromosome 3 clone RP11-479J2 map 3, WORKING DRAFT
             SEQUENCE, 26 unordered pieces.
ACCESSION    AC024447
VERSION      AC024447.2  GI:7534015
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 204993)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens chromosome 3, clone RP11-479J2
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 204993)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
             Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
             Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
             DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
             Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
             Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
             Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
             Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
             Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
             Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
             McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
             Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
             Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
             Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
             Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
             Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
             Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
             Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
             Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
             Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On Apr 11, 2000 this sequence version replaced gi:7108243.
             All repeats were identified using RepeatMasker:
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             -------------- Genome Center
                 Center: Whitehead Institute/ MIT Center for Genome Research
                 Center code: WIBR
                 Web site: http://www-seq.wi.mit.edu
                 Contact: sequence_submissions@genome.wi.mit.edu
             -------------- Project Information
                 Center project name: L6997
                 Center clone name: 479_J_2
             -------------- Summary Statistics
                 Sequencing vector: M13; M77815; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.960731
                 Consensus quality: 184405 bases at least Q40
                 Consensus quality: 195348 bases at least Q30
                 Consensus quality: 199562 bases at least Q20
                 Insert size: 202493; sum-of-contigs
                 Quality coverage: 3.5 in Q20 bases; sum-of-contigs
             --------------.
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 26 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *       1     1059: contig of 1059 bp in length
             *    1060  1159: gap of      100 bp
             *    1160     3098: contig of 1939 bp in length
             *    3099  3198: gap of      100 bp
             *    3199     4151: contig of 953 bp in length
             *    4152  4251: gap of      100 bp
             *    4252     6077: contig of 1826 bp in length
             *    6078  6177: gap of      100 bp
             *    6178    10158: contig of 3981 bp in length
             *   10159 10258: gap of      100 bp
             *   10259    13440: contig of 3182 bp in length
             *   13441 13540: gap of      100 bp
             *   13541    16429: contig of 2889 bp in length
             *   16430 16529: gap of      100 bp
             *   16530    19606: contig of 3077 bp in length
             *   19607 19706: gap of      100 bp
             *   19707    24259: contig of 4553 bp in length
             *   24260 24359: gap of      100 bp
             *   24360    28314: contig of 3955 bp in length
             *   28315 28414: gap of      100 bp
             *   28415    32584: contig of 4170 bp in length
             *   32585 32684: gap of      100 bp
             *   32685    37793: contig of 5109 bp in length
             *   37794 37893: gap of      100 bp
```

```
            *    37894    44505: contig of 6612 bp in length
            *    44506 44605: gap of      100 bp
            *    44606    50170: contig of 5565 bp in length
            *    50171 50270: gap of      100 bp
            *    50271    57212: contig of 6942 bp in length
            *    57213 57312: gap of      100 bp
            *    57313    68003: contig of 10691 bp in length
            *    68004 68103: gap of      100 bp
            *    68104    77484: contig of 9381 bp in length
            *    77485 77584: gap of      100 bp
            *    77585    86863: contig of 9279 bp in length
            *    86864 86963: gap of      100 bp
            *    86964    96495: contig of 9532 bp in length
            *    96496 96595: gap of      100 bp
            *    96596   107929: contig of 11334 bp in length
            *   107930 108029: gap of      100 bp
            *   108030   116856: contig of 8827 bp in length
            *   116857 116956: gap of      100 bp
            *   116957   128225: contig of 11269 bp in length
            *   128226 128325: gap of      100 bp
            *   128326   142566: contig of 14241 bp in length
            *   142567 142666: gap of      100 bp
            *   142667   160140: contig of 17474 bp in length
            *   160141 160240: gap of      100 bp
            *   160241   179301: contig of 19061 bp in length
            *   179302 179401: gap of      100 bp
            *   179402   204993: contig of 25592 bp in length.
FEATURES             Location/Qualifiers
     source          1. .204993
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-479J2"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1059
                     /note="assembly_fragment"
     misc_feature    1160. .3098
                     /note="assembly_fragment"
     misc_feature    3199. .4151
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    4252. .6077
                     /note="assembly_fragment"
     misc_feature    6178. .10158
                     /note="assembly_fragment"
     misc_feature    10259. .13440
                     /note="assembly_fragment"
     misc_feature    13541. .16429
                     /note="assembly_fragment"
     misc_feature    16530. .19606
                     /note="assembly_fragment"
     misc_feature    19707. .24259
                     /note="assembly_fragment"
     misc_feature    24360. .28314
                     /note="assembly_fragment"
     misc_feature    28415. .32584
                     /note="assembly_fragment"
     misc_feature    32685. .37793
                     /note="assembly_fragment"
     misc_feature    37894. .44505
                     /note="assembly_fragment"
     misc_feature    44606. .50170
                     /note="assembly_fragment"
     misc_feature    50271. .57212
                     /note="assembly_fragment"
     misc_feature    57313. .68003
                     /note="assembly_fragment"
     misc_feature    68104. .77484
                     /note="assembly_fragment"
     misc_feature    77585. .86863
                     /note="assembly_fragment"
     misc_feature    86964. .96495
                     /note="assembly_fragment"
     misc_feature    96596. .107929
                     /note="assembly_fragment"
     misc_feature    108030. .116856
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    116957. .128225
                     /note="assembly_fragment"
     misc_feature    128326. .142566
                     /note="assembly_fragment"
     misc_feature    142667. .160140
                     /note="assembly_fragment"
     misc_feature    160241. .179301
                     /note="assembly_fragment"
     misc_feature    179402. .204993
                     /note="assembly_fragment"
BASE COUNT     62968 a   39652 c   40375 g   59489 t    2509 others
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC024447    ..

 Align seg 1/1  to: AC024447  from: 1  to: 204993

       8 LeuThrIleSerAlaLeuLeuThrAla 16
         ||||||||||||||||||||||||||||
  180405 CTCACAATCTCTGCTCTCCTGACAGCA 180431

seq_name: gb_htg:AC092942

seq_documentation_block:
LOCUS        AC092942              205825 bp    DNA     linear   HTG 03-JAN-2002
DEFINITION   Homo sapiens chromosome 3q clone RP11-479J2, WORKING DRAFT
             SEQUENCE, 19 unordered pieces.
ACCESSION    AC092942
VERSION      AC092942.2  GI:18034707
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 205825)
  AUTHORS    Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
             Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
             Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
             Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
             Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
             Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
             Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
             Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
             Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
             Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
             Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
             Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
             Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
             Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
             Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
             Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
             Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
             Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
             Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
             Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
             Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
             Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
             Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
```

```
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 205825)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 2, 2002 this sequence version replaced gi:15135816.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HCZH
                Center clone name: RP11-479J2
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 198598 bases at least Q40
                Consensus quality: 208850 bases at least Q30
                Consensus quality: 214667 bases at least Q20
                Estimated insert size: 207320; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 3.5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    26477: contig of 26477 bp in length
            *    26478    26577: gap of unknown length
            *    26578    49296: contig of 22719 bp in length
            *    49297    49396: gap of unknown length
            *    49397    68898: contig of 19502 bp in length
            *    68899    68998: gap of unknown length
            *    68999    86550: contig of 17552 bp in length
            *    86551    86650: gap of unknown length
            *    86651   100899: contig of 14249 bp in length
            *   100900   100999: gap of unknown length
            *   101000   112284: contig of 11285 bp in length
            *   112285   112384: gap of unknown length
            *   112385   121682: contig of 9298 bp in length
            *   121683   121782: gap of unknown length
            *   121783   135703: contig of 13921 bp in length
            *   135704   135803: gap of unknown length
            *   135804   145468: contig of 9665 bp in length
            *   145469   145568: gap of unknown length
            *   145569   154983: contig of 9415 bp in length
            *   154984   155083: gap of unknown length
            *   155084   164549: contig of 9466 bp in length
            *   164550   164649: gap of unknown length
            *   164650   171823: contig of 7174 bp in length
            *   171824   171923: gap of unknown length
            *   171924   180207: contig of 8284 bp in length
            *   180208   180307: gap of unknown length
            *   180308   187008: contig of 6701 bp in length
            *   187009   187108: gap of unknown length
            *   187109   191344: contig of 4236 bp in length
            *   191345   191444: gap of unknown length
            *   191445   195679: contig of 4235 bp in length
            *   195680   195779: gap of unknown length
            *   195780   199345: contig of 3566 bp in length
            *   199346   199445: gap of unknown length
            *   199446   203512: contig of 4067 bp in length
            *   203513   203612: gap of unknown length
            *   203613   205825: contig of 2213 bp in length.
FEATURES             Location/Qualifiers
     source          1. .205825
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3q"
                     /clone="RP11-479J2"
BASE COUNT    63155 a  40250 c  40422 g  60165 t   1833 others
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC092942   ..

 Align seg 1/1  to: AC092942  from: 1  to: 205825

       8 LeuThrIleSerAlaLeuLeuThrAla 16
         ||||||||||||||||||||||||||||
    1010 CTCACAATCTCTGCTCTCCTGACAGCA 1036

seq_name: gb_htg:AC099581

seq_documentation_block:
LOCUS       AC099581             209518 bp    DNA     linear   HTG 16-NOV-2001
DEFINITION  Mus musculus chromosome 4 clone RP23-184J13 map 4, WORKING DRAFT
            SEQUENCE, 13 unordered pieces.
ACCESSION   AC099581
VERSION     AC099581.1  GI:16946030
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 209518)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 4, clone RP23-184J13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 209518)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
```

```
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L16751
                Center clone name: 184_J_13
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 206496 bases at least Q40
                Consensus quality: 207387 bases at least Q30
                Consensus quality: 207727 bases at least Q20
                Insert size: 214000; agarose-fp
                Insert size: 208318; sum-of-contigs
                Quality coverage: 8.8 in Q20 bases; agarose-fp
                Quality coverage: 9.1 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    12457: contig of 12457 bp in length
            *    12458 12557: gap of      100 bp
            *    12558    13208: contig of 651 bp in length
            *    13209 13308: gap of      100 bp
            *    13309    14078: contig of 770 bp in length
            *    14079 14178: gap of      100 bp
            *    14179    16566: contig of 2388 bp in length
            *    16567 16666: gap of      100 bp
            *    16667    20489: contig of 3823 bp in length
            *    20490 20589: gap of      100 bp
            *    20590    25395: contig of 4806 bp in length
            *    25396 25495: gap of      100 bp
            *    25496    32724: contig of 7229 bp in length
            *    32725 32824: gap of      100 bp
            *    32825    42342: contig of 9518 bp in length
            *    42343 42442: gap of      100 bp
            *    42443    49596: contig of 7154 bp in length
            *    49597 49696: gap of      100 bp
            *    49697    67429: contig of 17733 bp in length
            *    67430 67529: gap of      100 bp
            *    67530   102738: contig of 35209 bp in length
            *   102739 102838: gap of      100 bp
            *   102839   172277: contig of 69439 bp in length
            *   172278 172377: gap of      100 bp
            *   172378   209518: contig of 37141 bp in length.
FEATURES            Location/Qualifiers
     source         1. .209518
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /chromosome="4"
                    /map="4"
                    /clone="RP23-184J13"
                    /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature   1. .12457
                    /note="assembly_fragment
                    clone_end:SP6
                    vector_side:left"
     misc_feature   12558. .13208
                    /note="assembly_fragment"
     misc_feature   13309. .14078
                    /note="assembly_fragment"
     misc_feature   14179. .16566
                    /note="assembly_fragment"
     misc_feature   16667. .20489
                    /note="assembly_fragment"
     misc_feature   20590. .25395
                    /note="assembly_fragment"
     misc_feature   25496. .32724
                    /note="assembly_fragment"
     misc_feature   32825. .42342
                    /note="assembly_fragment"
     misc_feature   42443. .49596
                    /note="assembly_fragment"
     misc_feature   49697. .67429
                    /note="assembly_fragment"
     misc_feature   67530. .102738
                    /note="assembly_fragment"
     misc_feature   102839. .172277
                    /note="assembly_fragment"
     misc_feature   172378. .209518
                    /note="assembly_fragment
                    clone_end:T7
                    vector_side:right"
BASE COUNT    55725 a  44920 c  47112 g  60522 t   1239 others
ORIGIN


alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC099581   ..

 Align seg 1/1  to: AC099581  from: 1  to: 209518

      12 AlaLeuLeuThrAlaLeuLeuValThr 20
         ||||||||||||||||||||||||||||
    5884 GCCCTCCTCACAGCACTACTAGTTACA 5910

seq_name: gb_ba:AP003590

seq_documentation_block:
LOCUS       AP003590              333500 bp    DNA     linear   BCT 28-NOV-2001
DEFINITION  Nostoc sp. PCC 7120 DNA, complete genome, section 10/19.
ACCESSION   AP003590 BA000019
VERSION     AP003590.1  GI:17131676
KEYWORDS    .
SOURCE      Nostoc sp. PCC 7120 DNA.
  ORGANISM  Nostoc sp. PCC 7120
            Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE   1  (sites)
  AUTHORS   Kaneko,T., Nakamura,Y., Wolk,C.P., Kuritz,T., Sasamoto,S.,
            Watanabe,A., Iriguchi,M., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A.,
            Nakazaki,N., Shimpo,S., Sugimoto,M., Takazawa,M., Yamada,M.,
            Yasuda,M. and Tabata,S.
  TITLE     Complete genomic sequence of the filamentous nitrogen-fixing
            cyanobacterium Anabaena sp. strain PCC 7120
  JOURNAL   DNA Res. 8 (5), 205-213 (2001)
  MEDLINE   21595285
   PUBMED   11759840
REFERENCE   2  (bases 1 to 333500)
```

```
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/cyanobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1. .333500
                     /organism="Nostoc sp. PCC 7120"
                     /db_xref="taxon:103690"
                     /note="synonym:Anabaena sp. PCC7120"
     gene            complement(57. .926)
                     /gene="all2584"
     CDS             complement(57. .926)
                     /gene="all2584"
                     /note="ORF_ID:all2584"
                     /codon_start=1
                     /transl_table=11
                     /product="ATP-binding protein of ferrichrome ABC
                     transporter"
                     /protein_id="BAB74283.1"
                     /db_xref="GI:17131677"
                     /translation="MSVKPARLTLNSQHSTVNTQKIANMKGLSTQSLSLAYDGVPIIR
                     DLNLAIPTGKISALVGANGCGKSTLLRGLARLLKPYGGTVHLDGQSIFSLATKEVAKQ
                     LGILPQGPVAPEGLTVKDLVGQGRYPYQNWLQQWSEKDEIIVQQALEITNLLSLADRA
                     LDTLSGGQRQRAWIAMALAQDTDILLLDEPTTFLDLAHQIEVLDLLYELNQHQERTIV
                     MVLHDLNQACRYADYLVVVKQGRIFTAGEPKDVMTEEMVQEVFGLQCRIVSDPVVGTP
                     MCVPIGRKGEKQL"
     gene            complement(964. .2004)
                     /gene="all2585"
     CDS             complement(964. .2004)
                     /gene="all2585"
                     /note="ORF_ID:all2585"
                     /codon_start=1
                     /transl_table=11
                     /product="iron(III) dicitrate transport system permease
                     protein"
                     /protein_id="BAB74284.1"
                     /db_xref="GI:17131678"
                     /translation="MKVDWLVIRSETMSLRIDRRVPPILLCLVVAIVVAMVMNLGRGE
                     YPIAPLDIIKTVLGIDTGNPDHAFVIYNLRLPRTLVACMVGVALAVAGTIFQGITRNP
                     LADPGIIGINAGASLAAVTVIVLFPSAPIYTLPLSAFAGALLMAGLIYSLAWNNGSSP
                     VLFILMGVGLSAIAGAFTSLMITFGDIYSVSDALVWLAGSVYGRTWEQVFSFLPWLIV
                     FVPMALTLARHLNTLNLGDDVAKGLGTRVEWQRGLLVLVGVALAGAGVATAGMIGFVG
                     LIAPHLGRQLVGTNHQGLLPTSALLGGMLVVVADFLGRTLFAPIEIPCGVVTAAVGAP
                     YFLYLLIRNRKK"
     gene            complement(2001. .3032)
                     /gene="all2586"
     CDS             complement(2001. .3032)
                     /gene="all2586"
                     /note="ORF_ID:all2586"
                     /codon_start=1
                     /transl_table=11
                     /product="iron(III) dicitrate transport system permease
                     protein"
                     /protein_id="BAB74285.1"
                     /db_xref="GI:17131679"
                     /translation="MTRATTASPRNWNPPKISPLVGLILGILILLICLVYSVTLGAAE
                     IPLNKILESFITFDGSYEHLVIQTVRLPRSLVALLVGSSLAVAGALMQGLTRNPLADP
                     GILGIESGAALAVVTTIFVFGSSSLGLLTTVAFLGAGVTAMLVYFLGSLGKGGATPLN
                     LTVAGAALTALISSLTTAILIVSQRTLEEIRFWLAGSLAGRDFNILLSALPFVMIGLV
                     VAFALGRQITTMSLGEDMAKGLGQQTAWVKITTAISVVLLAGSSVSLAGPIGFIGLVV
                     PHMVRFFIKADYRWILPYSAVVGATLLLVADVAARVLLKPQELPVGVMTALVGAPFFV
                     YLAKSKVKK"
     gene            3507. .4490
                     /gene="alr2587"
     CDS             3507. .4490
                     /gene="alr2587"
                     /note="ORF_ID:alr2587"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator"
                     /protein_id="BAB74286.1"
                     /db_xref="GI:17131680"
                     /translation="MVLILQESADCYQSLTENNNTPQSNIWEAHIHDPQGMSHGYRQW
                     FHLRPGISLLTDHYQLNNDLIVEKISAPNCLWLELSFSIFGNNNNEDVYSGQNFLYVY
                     RELGKAGYTKWQAQERILKFDIHIEYSQIQKLLAHQMDVLPVSFRQMFEKNDDEPNYR
                     NFGITTHEMQTVLRQILNCPYQGLTKQLYLEGKILELLALRLEYLQENPNFNKSTTLK
                     PHQLDSIYHARDILINNINNPPSLINLARQVGLNDCTLKKGFQEIFGTTVFGYLHEYR
                     MEEAKELLLDKKMNVTQVAQIVGYEARTSFIRAFRKKFGVSPTAYVQKNSA"
     gene            4640. .7195
                     /gene="alr2588"
     CDS             4640. .7195
                     /gene="alr2588"
                     /note="ORF_ID:alr2588"
                     /codon_start=1
                     /transl_table=11
                     /product="ferrichrome-iron receptor"
                     /protein_id="BAB74287.1"
                     /db_xref="GI:17131681"
                     /translation="MRGEMLIKPALLHRLCLAATVSVLISNQTVWAEVNTSQNKTPQL
                     SKGVPILAQAPTTEILEVTGVKANPTEKGVEVILQTTSGEKLQLVNRSQGNAYIADIP
                     NAQLRLPNGDTFTFTSPKPVEGISEIIVSNLDANTIRVTVTGEAGLPVVELFDSDEGL
                     ILGVLSAAPSQTQATPTPEPEQPAATTDQPIELVVTGQQDSYRVTNAATATKTDTPLR
                     DIPQSIQVIPRQVIEDQRATQLNEALRNVSGVQPSNSAGRTRDRFVIRGFDDFNSVIR
                     DGFKENTSVYRETANIERIEVLKGPASVLFGQLEPGGVINVITKQPQREPFLIMGLEA
                     GSYGFFRPTVDFNSPLNDSKTLRLRVNAAVEISESFRDFDKETSRFFIAPVLAWEIGD
                     RTSLVFDLEYLKDSRPFDRGLVALGDRVAEIPFNRILGEPFDSKEVEDLRIGTRFEHG
                     FNDNWKLRSAFRIVSTQTSSFVTEPGSLDEATGLLSRDFGVDRPTPDETYAFQTDLIG
                     KFKTGAIEHELVIGFDFNRQTNFVDDRRSAPAPAIDIFNPVYLTARPNIESDPFRGLF
                     AADNIGIYIQDQIKLAENLKLLIGGRYDFTNQSFLSIVDGVKFFDISDQAFSPRIGIV
                     YQPITPLSLYASYSRSFAQNFGIQRDNSRIEPERGTQYEVGVRGEFLDGKLIASLAGY
                     QITKTNVATPDPADLNFSIPVGEVRSRGIEFDIAGELAKGWNIIASYAYTDAKITEDN
                     TDNEGNRLNRVPENSASIWTTYELQSGALQGLGMGVGLFFVGERQGDLSNSFTVPGYT
                     RTDAALFYRRDNWNIGLNFKNIFDVNYIESTTSRTSIDAGIPFTVIGSLSVKF"
     gene            7241. .8194
                     /gene="alr2589"
     CDS             7241. .8194
                     /gene="alr2589"
                     /note="ORF_ID:alr2589
                     hypothetical protein"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB74288.1"
                     /db_xref="GI:17131682"
                     /translation="MNYFFCADESRQVNEDPIGSGCRRQCYVLIECPPPWMANEFDSK
                     AVPANLRSLENEFYAVEPDGILLFIYSEELRQNYLTRILIFRPMMGYSGGYIKQEIHV
                     SSLEDVAIVLKDWLAGKDIADTDINPHTRDILVCTHGTHDKCCAKYGIPFYRQAKNIV
                     ADLAVESVRIWQSSHFSGHRFAPTILDLPQGRYYGKLDCQSFTNILTYTGDINCLKSI
                     YRGWGILPFPAQVMEGELIFLHGWQWFNYKVSCCFIEQSEDECFNRIEISAILPDGDI
                     KTYQADVVEDESKSIYLIGDCHGIEPELIPKYSVNNLIQII"
     gene            8195. .9187
                     /gene="alr2590"
     CDS             8195. .9187
                     /gene="alr2590"
                     /note="ORF_ID:alr2590"
                     /codon_start=1
                     /transl_table=11
                     /product="iron(III) dicitrate-binding periplasmic protein"
                     /protein_id="BAB74289.1"
                     /db_xref="GI:17131683"
                     /translation="MKTWFKYIKLLLLSVCTFLLVTGCNSNVSKLPSNQTTSNITTNC
                     RVINHEAGKTQICGQPKKVVALSPPVLDMMLALGVQPAGYAEVDLLNSKVFDRPKEQI
                     PYLGDRITSQPMNVGDRGNPSLESLLQLKPDLILGEASYNETEYQILNKIAPTILISH
                     RDGITPAWQQTISIIAQALGREDKVPTVIAEYEQKLSQAKTAIAPIAQHQELLLLAFR
                     GITLTSFTFGAETFAGGLLQDLGFKLVSPESTEYEVALEVLPKFKSDLIVIMPSGNNN
                     IENAKRQWSQNPILQSISAHNTNRIYFIDYQLASRIRGPIAAELFVNQVRQLLS"
     gene            9438. .10424
                     /gene="alr2591"
     CDS             9438. .10424
                     /gene="alr2591"
                     /note="ORF_ID:alr2591"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="transcriptional regulator"
                     /protein_id="BAB74290.1"
                     /db_xref="GI:17131684"
                     /translation="MVLILQESADCYQPLTENNNTPQSTIWEAHIHDPQGMSHGYRQW
                     FHLRPGISLLTDCYQLNDDLVVEKAPSPALIWLELSFNIFGHNICEGVYPGENFLDVY
                     RGLEPANHTEWQAQEKVLKFDIHLEYSELQRLFAHQMDLLPASFRRMFEKNDDEPEYR
                     NFGITTHEMQIVLRQILNCPYQGLTKQLYLEGKILELLALRLEYLQENPTNINTDNTL
                     KNHQVDGIYHARDILIKNLNNPPSLIDLARLVGLNDCTLKKGFQEVFGTTAFGYLHEY
                     RMEQAKELLLEKEMNVTQVAQVVGYEARTSFIRAFRKKFGVSPTAYIQRNSA"
     gene            10585. .13134
                     /gene="alr2592"
     CDS             10585. .13134
                     /gene="alr2592"
                     /note="ORF_ID:alr2592"
                     /codon_start=1
                     /transl_table=11
                     /product="ferrichrome-iron receptor"
                     /protein_id="BAB74291.1"
                     /db_xref="GI:17131685"
                     /translation="MLIKAGLLNTLYLATIVSILISNQTVRAEVNTTGNQTPQLSKDV
                     SILAQAPTTEIVQVTGVTANPTEKGVEVILQTTSGEKLQLVNLSQGNAYIVDIPNAQL
                     RLPNGDSFTFTSPKPVEGISEIIVTNLDANTIRVTVTGETGLPVVELFDSDEGLILGV
                     VSATPPQTQATPEQPAAQTEENQEPIELVVTGEQDIYSVPNASTATRTDTPLRDIPQS
                     IQVIPQQVLKDQQVTRILDAVRNVSGVTPQKGYGDATDFYTIRGFNSNRTLRNGFNVR
                     AARGGTSTFTAPSSVERVEVLKGPASVLYGQLEPGGLVNIVTKKPLSDPYYAAEFTAG
                     SYSFYSPSLDISGPLTTDKKLLYRLNASYQNFGSFIDFVNGENLSIVPVVNYQISDRT
                     ALTFEYEHSRSNQSFNDGLPIDPISFELPISRNLNNPDDTLDNTTNSLNFTLEHGFND
                     NIRLRTVFGASFSEGTTLAYRLFNYNPETNEVLRAFRNSPDYENDYSWQTDVIAKFNT
                     GAVKHELLVGFELAFNDNAYNATDSEDSFINVFNPVYDTPIPLALYSFSGKSNSNTIG
                     LYLQDQITLSPNLKLLIGGRYDFARAKSDSNFQFIGEPDSPTSDEFYNEAFSPRVGLV
                     YQPIEPISLYASYSRSFLPNNAKTFSGEIIEPTRGKQLEVGVKADFGRLSTTLAAYEI
                     TKTNILTSDRDNPFFSLGVGEVRSRGIELDISGEILPGWNVIASTFINDSVITQDNNL
                     PIGDTLVNAPRRGASLWTTYEIQTGDLQGLGFGLGVFYVGDRQAELPNNLVLPSYVRA
                     DASIFYRRDNWRIGLNFKNLSDTRYYESSQNSSLIYPGTPLTVLGTVSVQF"

alignment_scores:
            Quality:    9.00              Length:      9
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AP003590/rev   ..

 Align seg 1/1  to reverse of: AP003590  from: 1  to: 333500

      85 IleLeuSerLeuIleIleSerPheLeu 93
         |||||||||||||||||||||||||||
  276344 ATATTATCTTTAATTATCTCATTCCTA 276318

seq_name: gb_pat:AX120599

seq_documentation_block:
LOCUS       AX120599                 261 bp    DNA     linear   PAT 11-MAY-2001
DEFINITION  Sequence 515 from Patent EP1108790.
ACCESSION   AX120599
VERSION     AX120599.1  GI:14037314
KEYWORDS    .
SOURCE      Corynebacterium glutamicum.
  ORGANISM  Corynebacterium glutamicum
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Corynebacteriaceae;
            Corynebacterium.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
            Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
  TITLE     Novel polynucleotides
  JOURNAL   Patent: EP 1108790-A 515 20-JUN-2001;
            KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES             Location/Qualifiers
     source          1. .261
                     /organism="Corynebacterium glutamicum"
                     /db_xref="taxon:1718"
BASE COUNT       52 a     77 c     54 g     78 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AX120599   ..

 Align seg 1/1  to: AX120599  from: 1  to: 261

      11 SerAlaLeuLeuThrAlaLeuLeu 18
         ||||||||||||||||||||||||
      95 TCCGCTTTGCTCACAGCCCTTCTT 118

seq_name: gb_pl:AF314550

seq_documentation_block:
LOCUS       AF314550                 293 bp    mRNA    linear   PLN 02-NOV-2001
DEFINITION  Glycine max clone NTN5c NBS-type putative resistance protein mRNA,
            partial cds.
ACCESSION   AF314550
VERSION     AF314550.1  GI:16588591
KEYWORDS    .
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 293)
  AUTHORS   Jeong,S.C., Hayes,A.J., Biyashev,R.M. and Saghai Maroof,M.A.
  TITLE     Diversity and evolution of a non-TIR-NBS sequence family that
            clusters to a chromosomal 'hotspot' for disease resistance genes in
            soybean
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 293)
  AUTHORS   Jeong,S.C., Hayes,A.J., Biyashev,R.M. and Saghai Maroof,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-2000) Crop and Soil Environmental Sciences,
            Virginia Tech, 350 Smyth Hall, Blacksburg, VA 24061-0404, USA
FEATURES             Location/Qualifiers
     source          1. .293
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="NTN5c"
     CDS             <1. .>293
                     /codon_start=2
                     /product="NBS-type putative resistance protein"
                     /protein_id="AAL26851.1"
                     /db_xref="GI:16588592"
                     /translation="MGGVGKTTLAQHVYNDPRIEGKFDIKAWVCVSDDFDVLTVTRAI
                     LEAVIDSTDNSRGLEMVHRRLKENLIGKRFLLVLDDVWNEKREKWEAVLTPLN"
BASE COUNT       89 a     41 c     79 g     84 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AF314550   ..

 Align seg 1/1  to: AF314550  from: 1  to: 293

      50 GlnGluIleSerSerArgPheGly 57
         |||||||||||||||||||||||
     214 CAAGAGATTTCTTCTCGTTTTGGA 237
```

```
seq_name: gb_pl:AF222879

seq_documentation_block:
LOCUS       AF222879               312 bp    DNA     linear   PLN 15-MAR-2000
DEFINITION  Glycine max clone R14 disease resistance-like protein gene, partial
            cds.
ACCESSION   AF222879
VERSION     AF222879.1  GI:7243628
KEYWORDS    .
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 312)
  AUTHORS   Hayes,A.J. and Saghai Maroof,M.A.
  TITLE     Targeted resistance gene mapping in soybean using modified AFLPs
  JOURNAL   Theor. Appl. Genet. (2000) In press
REFERENCE   2  (bases 1 to 312)
  AUTHORS   Hayes,A.J. and Saghai Maroof,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2000) Crop and Soil Environmental Sciences,
            Virginia Tech, Blacksburg, VA 24061-0404, USA
FEATURES             Location/Qualifiers
     source          1. .312
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="R14"
     CDS             <1. .>312
                     /codon_start=1
                     /product="disease resistance-like protein"
                     /protein_id="AAF43387.1"
                     /db_xref="GI:7243629"
                     /translation="GMGGVGKTTLAQHVYNDPRIEGKFDIKAWVCVSDDFDVLTVTRA
                     ILEAVIDSTDNSRGLEMVHRRLKENLIGKRFLLVLDDVWNEKREEWEAVQTPLTYGAR
                     GS"
BASE COUNT       90 a     46 c     89 g     87 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AF222879   ..

 Align seg 1/1  to: AF222879  from: 1  to: 312

      50 GlnGluIleSerSerArgPheGly 57
         |||||||||||||||||||||||
     216 CAAGAGATTTCTTCTCGTTTTGGA 239

seq_name: gb_in:TIU42193

seq_documentation_block:
LOCUS       TIU42193               407 bp    DNA     linear   INV 13-MAR-1996
DEFINITION  Taeniothrips inconsequens cytochrome oxidase I (CO1) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U42193
VERSION     U42193.1  GI:1223995
KEYWORDS    .
SOURCE      Taeniothrips inconsequens.
  ORGANISM  Mitochondrion Taeniothrips inconsequens
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Paraneoptera; Thysanoptera; Terebrantia;
            Thripoidea; Thripidae; Taeniothrips.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Crespi,B.J., Carmean,D.A., Vawter,L. and von Dohlen,C.
  TITLE     Molecular Phylogenetics of Thysanoptera
  JOURNAL   Systematic Entomology (1996) In press
REFERENCE   2  (bases 1 to 407)
  AUTHORS   Carmean,D.A., Crespi,B.J., Vawter,L. and von Dohlen,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-1995) David A. Carmean, Biological Sciences,
            Simon Fraser University, Burnaby, BC V5A1S6, Canada
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Taeniothrips inconsequens"
                     /organelle="mitochondrion"
                     /db_xref="taxon:45055"
     gene            1. .407
                     /gene="CO1"
     CDS             <1. .>407
                     /gene="CO1"
                     /codon_start=3
                     /transl_table=5
                     /product="cytochrome oxidase I"
                     /protein_id="AAA92103.1"
                     /db_xref="GI:1223996"
                     /translation="LPPSIILLIMGLSKEGAGTGWTVYPPLSTFYHSGISVDLTIFSL
                     HLAGISSILGALNFITTIMNLKIKNLSMSKTTLFVWSVILTAILLLLSLPVLAGAITM
                     LLTDRNLNTSFFDPGGGGDPVLYQHLFWFFGHP"
BASE COUNT      126 a     75 c     52 g    154 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x TIU42193/rev   ..

 Align seg 1/1  to reverse of: TIU42193  from: 1  to: 407

      64 PheMetIleValValIleLysPhe 71
         |||||||||||||||||||||||
     193 TTTATGATTGTAGTAATAAAATTT 170

seq_name: gb_pl:AF060192

seq_documentation_block:
LOCUS       AF060192               519 bp    DNA     linear   PLN 23-FEB-2001
DEFINITION  Glycine max putative resistance protein KNBS4 gene, partial cds.
ACCESSION   AF060192 AF325687
VERSION     AF060192.2  GI:13111696
KEYWORDS    .
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 519)
  AUTHORS   He,C. and Chen,S.Y.
  TITLE     Resistance gene analogs from soybean
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 519)
  AUTHORS   Dong,W. and Chen,S.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-1998) 803 Lab, Institute of Genetics, CAS,
            Beijing 100101, P.R. China
REFERENCE   3  (bases 1 to 519)
  AUTHORS   He,C. and Chen,S.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-FEB-2001) Plant Biotechnology Laboratory, Institute
            of Genetics, CAS, Datun Road, Beijing 100101, P.R. China
  REMARK    Sequence update by submitter
COMMENT     On Feb 23, 2001 this sequence version replaced gi:3091215.
FEATURES             Location/Qualifiers
     source          1. .519
                     /organism="Glycine max"
```

```
                     /db_xref="taxon:3847"
     mRNA            <1. .>519
     CDS             /product="putative resistance protein KNBS4"
                     <1. .>519
                     /codon_start=1
                     /product="putative resistance protein KNBS4"
                     /protein_id="AAC15224.2"
                     /db_xref="GI:13111697"
                     /translation="SGGVGKTTLAQHVYSDPRIEGKFVIKAWVCVSDDFDVLTVTRAI
                     LEAVIDSTDNSRGLEMVHRRLKENLIGKRFLLVLDDVWNEKREKWEAVLTPLTYGARG
                     SRILVTTRTTKVASTVRSNKELHLEQLQEDHCWKVFAKHAFQDDNPRLNVELKEIGIK
                     IVEKCKGLPLALG"
BASE COUNT      155 a     79 c    139 g    146 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AF060192   ..

 Align seg 1/1  to: AF060192  from: 1  to: 519

      50 GlnGluIleSerSerArgPheGly 57
         |||||||||||||||||||||||
     213 CAAGAGATTTCTTCTCGTTTTGGA 236

seq_name: gb_pl:AB027824

seq_documentation_block:
LOCUS       AB027824                 521 bp    DNA     linear   PLN 21-JUN-2001
DEFINITION  Schizosaccharomyces pombe gene for Protein kinase C-like 2, partial
            cds, clone:S644.
ACCESSION   AB027824
VERSION     AB027824.1  GI:6478831
KEYWORDS    .
SOURCE      Schizosaccharomyces pombe (strain:968 h90) DNA, clone:S644.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
            Schizosaccharomycetales; Schizosaccharomycetaceae;
            Schizosaccharomyces.
REFERENCE   1  (sites)
  AUTHORS   Ding,D.Q., Tomita,Y., Yamamoto,A., Chikashige,Y., Haraguchi,T. and
            Hiraoka,Y.
  TITLE     Large-scale screening of intracellular protein localization in
            living fission yeast cells by the use of a GFP-fusion genomic DNA
            library
  JOURNAL   Genes Cells 5 (3), 169-190 (2000)
  MEDLINE   20223868
REFERENCE   2  (bases 1 to 521)
  AUTHORS   Ding,D., Tomita,Y. and Hiraoka,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-1999) Da-Qiao Ding, Communications Research
            Laboratory, Kansai Advanced Research Center; 588-2, Iwaoka,
            Iwaoka-cho, Nishiku, Kobe, Hyogo 651-2401, Japan
            (E-mail:ding@crl.go.jp, Tel:+81-78-969-2240, Fax:+81-78-969-2249)
COMMENT     An S. pombe gene library in which genomic DNA fragments are fused
            to the 5'-end of the GFP-S65T gene was constructed. S. pombe strain
            was transformed with the library DNA amplified in E.coli. Plasmids
            of those transformants that showed interesting GFP localization
            were isolated, and their nucleotide sequences were determined.
            Detailed information and images can be seen in our web site
            (http://www-karc.crl.go.jp/bio/CellMagic/). Note: Only major
            localization is described for each clone in most cases. Because it
            is a random fusion of DNA fragment with the GFP gene, the
            subcellular localization of the fusion protein may differ from that
            of the full-length gene product.
FEATURES             Location/Qualifiers
     source          1. .521
                     /organism="Schizosaccharomyces pombe"
                     /strain="968 h90"
                     /db_xref="taxon:4896"
                     /clone="S644"
                     /note="subcellular localization of GFP fusion; Cell
                     periphary and septum"
     gene            1. .521
                     /gene="pck2+"
     CDS             <1. .>521
                     /gene="pck2+"
                     /codon_start=1
                     /product="Protein kinase C-like 2"
                     /protein_id="BAA87128.1"
                     /db_xref="GI:6478832"
                     /translation="TETRIYFVMDFVSGGDLMLHIQQEQFSRRRAQFYAAEVCLALKY
                     FHDNGIIYRDLKLDNILLSPDGHVKVADYGLCKEDMWHDNTTATFCGTPEFMAPEILL
                     EQQYTRSVDWWAFGVLIYQMLLGQSPFRGEDEEEIFDAILSDEPLYPIHMPRDSVSIL
                     QQLLTRDPKKRLGS"
BASE COUNT      146 a     93 c    111 g    171 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AB027824   ..

 Align seg 1/1  to: AB027824  from: 1  to: 521

      85 IleLeuSerLeuIleIleSerPhe 92
         |||||||||||||||||||||||
     161 ATCTTAAGCTTGATAATATCCTTT 184

seq_name: gb_pat:AX317918

seq_documentation_block:
LOCUS       AX317918                 562 bp    DNA     linear   PAT 14-DEC-2001
DEFINITION  Sequence 19 from Patent WO0190149.
ACCESSION   AX317918
VERSION     AX317918.1  GI:17900772
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Vogeli,G.
  TITLE     G protein-coupled receptors
  JOURNAL   Patent: WO 0190149-A 19 29-NOV-2001;
            PHARMACIA & UPJOHN COMPANY (US)
FEATURES             Location/Qualifiers
     source          1. .562
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      176 a    118 c    140 g    128 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AX317918/rev   ..

 Align seg 1/1  to reverse of: AX317918  from: 1  to: 562

      10 IleSerAlaLeuLeuThrAlaLeu 17
         |||||||||||||||||||||||
      66 ATTTCTGCGCTTCTCACTGCCTTG 43
```

```
seq_name: gb_pl:AY063021

seq_documentation_block:
LOCUS       AY063021               847 bp    mRNA    linear   PLN 26-NOV-2001
DEFINITION  Arabidopsis thaliana putative myo-inositol monophosphatase
            (F13E7_19/AT3g02870) mRNA, complete cds.
ACCESSION   AY063021
VERSION     AY063021.1  GI:17104612
KEYWORDS    FLI_CDNA.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 847)
  AUTHORS   Yamada,K., Banh,J., Banno,F., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C., Toriumi,M., Wu,H.C.,
            Yamamura,Y., Yu,G., Yu,S., Bowser,L., Carninci,P., Chen,H.,
            Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Kim,C., Koesema,E., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tracy,S.E.,
            Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Full Length cDNA of gene F13E7_19/AT3g02870 (GI:6728975)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 847)
  AUTHORS   Yamada,K., Banh,J., Banno,F., Chang,E., Dale,J.M., Goldsmith,A.D.,
            Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamamura,Y., Yu,G., Yu,S., Bowser,L., Carninci,P., Chen,H.,
            Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A.,
            Karlin-Neumann,G., Kawai,J., Kim,C., Koesema,E., Lam,B., Lin,J.,
            Meyers,M.C., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J.,
            Sakurai,T., Satou,M., Seki,M., Shinn,P., Southwick,A., Tracy,S.E.,
            Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Plant Gene Expression Center, 800 Buchanan
            Street, Albany, CA 94710, USA
COMMENT     The RIKEN Genomic Sciences Center (GSC) members carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
            Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai, J.,
            Hayashizaki,Y. and Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members constructed and
            sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada,K.,
            Banh,J., Banno,F., Chang,E., Dale,J.M., Goldsmith,A.D., Lee,J.M.,
            Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wu,H.C.,
            Yamamura,Y., Yu,G., Yu,S., Bowser,L., Chen,H., Cheuk,R., Jones,T.,
            Karlin-Neumann,G., Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C.,
            Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A.,
            Tracy,S.E., Davis,R.W., Ecker,J.R. and Theologis,A.

            Yamada,K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally
            to this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP
            /PGEC) contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .847
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="U09692"
                     /note="This clone is in pUNI 51.
                     ecotype: Columbia"
     gene            1. .847
                     /gene="F13E7_19/AT3g02870"
     CDS             1. .816
                     /gene="F13E7_19/AT3g02870"
                     /codon_start=1
                     /evidence=experimental
                     /product="putative myo-inositol monophosphatase"
                     /protein_id="AAL34195.1"
                     /db_xref="GI:17104613"
                     /translation="MADNDSLDQFLAAAIDAAKKAGQIIRKGFYETKHVEHKGQVDLV
                     TETDKGCEELVFNHLKQLFPNHKFIGEETTAAFGVTELTDEPTWIVDPLDGTTNFVHG
                     FPFVCVSIGLTIGKVPVVGVVYNPIMEELFTGVQGKGAFLNGKRIKVSAQSELLTALL
                     VTEAGTKRDKATLDDTTNRINSLLTKVRSLRMSGSCALDLCGVACGRVDIFYELGFGG
                     PWDIAAGIVIVKEAGGLIFDPSGKDLDITSQRIAASNASLKELFAEALRLTGA"
     3'UTR           817. .847
                     /gene="F13E7_19/AT3g02870"
BASE COUNT      231 a    164 c    215 g    237 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AY063021   ..

 Align seg 1/1  to: AY063021  from: 1  to: 847

      13 LeuLeuThrAlaLeuLeuValThr 20
         ||||||||||||||||||||||||
     463 CTTTTAACCGCTTTGCTCGTGACA 486

seq_name: gb_in:HAY12271

seq_documentation_block:
LOCUS       HAY12271               867 bp    mRNA    linear   INV 29-MAY-1998
DEFINITION  H.armigera mRNA for putative serine protease, clone HaTS7.
ACCESSION   Y12271
VERSION     Y12271.1  GI:2463059
KEYWORDS    putative; serine protease; trypsin-like protease.
SOURCE      cotton bollworm.
  ORGANISM  Helicoverpa armigera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
            Ditrysia; Noctuoidea; Noctuidae; Heliothinae; Helicoverpa.
REFERENCE   1  (bases 1 to 867)
  AUTHORS   Bown,D.P., Wilkinson,H.S. and Gatehouse,J.A.
  TITLE     Differentially regulated inhibitor-sensitive and insensitive
            protease genes from the phytophagous insect pest, Helicoverpa
            armigera, are members of complex multigene families
  JOURNAL   Insect Biochem. Mol. Biol. 27 (7), 625-638 (1997)
  MEDLINE   98067794
REFERENCE   2  (bases 1 to 867)
  AUTHORS   Bown,D.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-1997) D.P. Bown, University Of Durham, Biological
            Sciences, Science Laboratories, South Road, Durham, DH1 3LE, UK
FEATURES             Location/Qualifiers
     source          1. .867
                     /organism="Helicoverpa armigera"
                     /db_xref="taxon:29058"
                     /clone="HaTS7"
                     /tissue_type="midgut"
                     /dev_stage="fourth instar larvae"
     sig_peptide     8. .55
     CDS             8. .772
                     /codon_start=1
                     /product="trypsin-like protease"
                     /protein_id="CAA72950.1"
                     /db_xref="GI:2463060"
                     /db_xref="SPTREMBL:O18436"
                     /translation="MRILALVALCFAAVAAVPSNPQRIVGGSVTTIDQYPTIAALLYS
                     WNLSTYWQACGGTILNNRAILTAAHCTAGDANNRWRIRLGSTWANSGGVVHNLNANIV
                     HPSYNSRTMDNDIAVLRSATTFSFNNQVRAASIAGANYNLADNQAVWAAGWGTTSSGG
                     SSSEQLRHVQLVTINQNTCRNNYATRGIAITDNMLCSGWPNGGRDQCQGDSGGPLYHN
                     GIVVGVCSFGIGCAQAAFPGVNARVSRYTSWISSNA"
BASE COUNT      207 a    268 c    189 g    203 t
ORIGIN
```

```
alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x HAY12271    ..

 Align seg 1/1  to: HAY12271  from: 1  to: 867

      13 LeuLeuThrAlaLeuLeuValThr 20
         ||||||||||||||||||||||||
     207 CTGCTCACTGCACTGCTGGTGACG 230

seq_name: gb_in:HAY12269

seq_documentation_block:
LOCUS       HAY12269                870 bp    mRNA    linear   INV 29-MAY-1998
DEFINITION  H.armigera mRNA for putative serine protease, clone HaTC11, HaTC21,
            HaTS4, SR99.
ACCESSION   Y12269
VERSION     Y12269.1  GI:2463055
KEYWORDS    putative; serine protease; trypsin-like protease.
SOURCE      cotton bollworm.
  ORGANISM  Helicoverpa armigera
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
            Ditrysia; Noctuoidea; Noctuidae; Heliothinae; Helicoverpa.
REFERENCE   1  (bases 1 to 870)
  AUTHORS   Bown,D.P., Wilkinson,H.S. and Gatehouse,J.A.
  TITLE     Differentially regulated inhibitor-sensitive and insensitive
            protease genes from the phytophagous insect pest, Helicoverpa
            armigera, are members of complex multigene families
  JOURNAL   Insect Biochem. Mol. Biol. 27 (7), 625-638 (1997)
  MEDLINE   98067794
REFERENCE   2  (bases 1 to 870)
  AUTHORS   Bown,D.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-1997) D.P. Bown, University Of Durham, Biological
            Sciences, Science Laboratories, South Road, Durham, DH1 3LE, UK
FEATURES             Location/Qualifiers
     source          1. .870
                     /organism="Helicoverpa armigera"
                     /db_xref="taxon:29058"
                     /clone="SR99"
                     /clone="HaTC11"
                     /clone="HaTC21"
                     /clone="HaTS4"
                     /tissue_type="midgut"
                     /dev_stage="fourth instar larvae"
     sig_peptide     12. .59
     CDS             12. .776
                     /codon_start=1
                     /product="trypsin-like protease"
                     /protein_id="CAA72948.1"
                     /db_xref="GI:2463056"
                     /db_xref="SPTREMBL:O18434"
                     /translation="MRILALVALCFAAVAAVPSNPQRIVGGSVTTIDQYPTIAALLYS
                     WNLSTYWQACGGTILNNRAILTAAHCTAGDANNRWRIRLGSTWANSGGVVHNLNANII
                     HPSYNSRTMDNDIAVLRSATTFSFNNQVRAASIAGANYNLADNQAVWAAGWGTTSSGG
                     SSSEQLRHVQLVTINQNTCRNNYATRGIAITDNMLCSGWPNGGRDQCQGDSGGPLYHN
                     GIVVGVCSFGIGCAQAAFPGVNARVSRYTSWISSNA"
     mat_peptide     81. .773
                     /product="trypsin-like protease"
BASE COUNT      208 a    265 c    191 g    206 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x HAY12269    ..

 Align seg 1/1  to: HAY12269  from: 1  to: 870

      13 LeuLeuThrAlaLeuLeuValThr 20
         ||||||||||||||||||||||||
     211 CTGCTCACTGCACTGCTGGTGACG 234

seq_name: gb_pat:AR020714

seq_documentation_block:
LOCUS       AR020714                885 bp    DNA     linear   PAT 05-DEC-1998
DEFINITION  Sequence 3 from patent US 5789193.
ACCESSION   AR020714
VERSION     AR020714.1  GI:3975329
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 885)
  AUTHORS   Keranen,S., Aalto,M., Outola,M., Ronne,H. and Penttila,M.
  TITLE     Increased production of secreted proteins by recombinant eukaryotic
            cells
  JOURNAL   Patent: US 5789193-A 3 04-AUG-1998;
FEATURES             Location/Qualifiers
     source          1. .885
                     /organism="unknown"
BASE COUNT      320 a    171 c    196 g    198 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AR020714/rev    ..

 Align seg 1/1  to reverse of: AR020714  from: 1  to: 885

      12 AlaLeuLeuThrAlaLeuLeuVal 19
         ||||||||||||||||||||||||
     779 GCACTCTTAACGGCCTTGTTGGTG 756

seq_name: gb_sts:CNS06FCD

seq_documentation_block:
LOCUS       CNS06FCD                972 bp    DNA     linear   STS 10-JAN-2001
DEFINITION  T7 end of clone AR0AA029A04 of library AR0AA from strain CBS 732 of
            Zygosaccharomyces rouxii, sequence tagged site.
ACCESSION   AL396227
VERSION     AL396227.1  GI:12148185
KEYWORDS    STS.
SOURCE      Zygosaccharomyces rouxii.
  ORGANISM  Zygosaccharomyces rouxii
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.
REFERENCE   1  (bases 1 to 972)
  AUTHORS   de Montigny,J., Straub,M., Potier,S., Tekaia,F., Dujon,B.,
            Wincker,P., Artiguenave,F. and Souciet,J.
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 8.
            Zygosaccharomyces rouxii(1)
  JOURNAL   FEBS Lett. 487 (1), 52-55 (2000)
  PUBMED    11152883
REFERENCE   2  (bases 1 to 972)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
```

```
  TITLE     Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
            yeast species for molecular evolution studies(1)
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
   PUBMED   11152876
REFERENCE   3  (bases 1 to 972)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This STS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .972
                     /organism="Zygosaccharomyces rouxii"
                     /strain="CBS 732"
                     /db_xref="taxon:4956"
                     /clone="AR0AA029A04"
                     /clone_lib="AR0AA"
                     /note="end : T7"
     STS             1. .972
BASE COUNT      345 a    166 c    240 g    218 t      3 others
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x CNS06FCD/rev   ..

 Align seg 1/1  to reverse of: CNS06FCD  from: 1  to: 972

      86 LeuSerLeuIleIleSerPheLeu 93
         |||||||||||||||||||||||||
     213 CTTTCCCTGATTATCTCTTTCTTG 190

seq_name: em_htgo_inv:AC043561

seq_documentation_block:
ID   AC043561   standard; DNA; HTG; 1005 BP.
XX
AC   AC043561;
XX
SV   AC043561.1
XX
DT   12-APR-2000 (Rel. 63, Created)
DT   12-APR-2000 (Rel. 63, Last updated, Version 1)
XX
DE   Giardia intestinalis clone EJ5353 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX
KW   HTG; HTGS_PHASE0.
XX
OS   Giardia intestinalis
OC   Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
XX
RN   [1]
RP   1-1005
RA   Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
RT   "Giardia: a model for ancient eukaryotic genome analysis";
RL   Unpublished.
XX
RN   [2]
RP   1-1005
RA   Eakin N.Q., Morrison H.G., McArthur A.G., Nixon J., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
RT   ;
RL   Submitted (11-APR-2000) to the EMBL/GenBank/DDBJ databases.
RL   Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL   Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC   * NOTE: This record contains 1 individual
CC   * sequencing reads that have not been assembled into
CC   * contigs. Runs of N are used to separate the reads
CC   * and the order in which they appear is completely
CC   * arbitrary. Low-pass sequence sampling is useful for
CC   * identifying clones that may be gene-rich and allows
CC   * overlap relationships among clones to be deduced.
CC   * However, it should not be assumed that this clone
CC   * will be sequenced to completion. In the event that
CC   * the record is updated, the accession number will
CC   * be preserved.
CC   *        1     1005: contig of 1005 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .1005
FT                   /db_xref="taxon:5741"
FT                   /organism="Giardia intestinalis"
FT                   /strain="WB-C6"
FT                   /clone="EJ5353"
XX
SQ   Sequence 1005 BP; 226 A; 266 C; 248 G; 265 T; 0 other;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC043561   ..

 Align seg 1/1  to: AC043561  from: 1  to: 1005

       7 ArgLeuThrIleSerAlaLeuLeu 14
         |||||||||||||||||||||||||
      74 CGCTTGACAATCTCGGCTCTTCTG 97

seq_name: gb_pl:AY035150

seq_documentation_block:
LOCUS       AY035150               1038 bp    mRNA    linear   PLN 10-JUN-2001
DEFINITION  Arabidopsis thaliana putative myo-inositol monophosphatase
            (F13E7_19/AT3g02870) mRNA, complete cds.
ACCESSION   AY035150
VERSION     AY035150.1  GI:14334953
KEYWORDS    FLI_CDNA.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1038)
  AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
            Goldsmith,A.D., Lee,J.M., Quach,H.L., Tang,C., Toriumi,M., Yu,G.,
            Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
            Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
            Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,
            Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
            Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and
            Theologis,A.
  TITLE     Full Length cDNA of gene F13E7_19/AT3g02870 (GI:6728975)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1038)
```

```
AUTHORS   Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
          Goldsmith,A.D., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G.,
          Bowser,L., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
          Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
          Kim,C., Koesema,E., Lam,B., Lin,J., Meyers,M.C., Miranda,M.,
          Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M.,
          Shinn,P., Southwick,A., Shinozaki,K., Davis,R.W., Ecker,J.R. and
          Theologis,A.
TITLE     Direct Submission
JOURNAL   Submitted (10-MAY-2001) Plant Gene Expression Center, 800 Buchanan
          Street, Albany, CA 94710, USA
COMMENT   RIKEN Genomic Sciences Center (GSC) members carried out the
          collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
          Arabidopsis Full-Length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
          Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
          Hayashizaki,Y. and Shinozaki,K.

          The Salk, Stanford, PGEC (SSP) Consortium members carried out the
          sequencing and annotation of the RAFL cDNAs: Yamada,K., Liu,S.X.,
          Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D.,
          Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Yu,G., Bowser,L.,
          Chen,H., Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Koesema,E.,
          Lam,B., Lin,J., Meyers,M.C., Miranda,M., Nguyen,M., Palm,C.J.,
          Shinn,P., Southwick,A., Davis,R.W., Ecker,J.R. and Theologis,A.

          Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
          this work.  Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
          contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .1038
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="3"
                     /clone="RAFL04-13-N21 (R09692)"
                     /note="This clone is in a modified pBluescript vector
                     (Lambda ZAP) as a XhoI/SstI insert.
                     ecotype: Columbia"
     gene            1. .1038
                     /gene="F13E7_19/AT3g02870"
     5'UTR           1. .88
                     /gene="F13E7_19/AT3g02870"
     CDS             89. .904
                     /gene="F13E7_19/AT3g02870"
                     /codon_start=1
                     /evidence=experimental
                     /product="putative myo-inositol monophosphatase"
                     /protein_id="AAK59654.1"
                     /db_xref="GI:14334954"
                     /translation="MADNDSLDQFLAAAIDAAKKAGQIIRKGFYETKHVEHKGQVDLV
                     TETDKGCEELVFNHLKQLFPNHKFIGEETTAAFGVTELTDEPTWIVDPLDGTTNFVHG
                     FPFVCVSIGLTIGKVPVVGVVYNPIMEELFTGVQGKGAFLNGKRIKVSAQSELLTALL
                     VTEAGTKRDKATLDDTTNRINSLLTKVRSLRMSGSCALDLCGVACGRVDIFYELGFGG
                     PWDIAAGIVIVKEAGGLIFDPSGKDLDITSQRIAASNASLKELFAEALRLTGA"
     3'UTR           905. .1038
                     /gene="F13E7_19/AT3g02870"
BASE COUNT      312 a    202 c    234 g    290 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AY035150   ..

 Align seg 1/1  to: AY035150  from: 1  to: 1038

      13 LeuLeuThrAlaLeuLeuValThr 20
         ||||||||||||||||||||||||
     551 CTTTTAACCGCTTTGCTCGTGACA 574
```

```
seq_name: em_htgo_inv:AC046945

seq_documentation_block:
ID   AC046945   standard; DNA; HTG; 1086 BP.
XX
AC   AC046945;
XX
SV   AC046945.1
XX
DT   14-APR-2000 (Rel. 63, Created)
DT   14-APR-2000 (Rel. 63, Last updated, Version 1)
XX
DE   Giardia intestinalis clone KI0649 strain WB-C6, LOW-PASS SEQUENCE SAMPLING.
XX
KW   HTG; HTGS_PHASE0.
XX
OS   Giardia intestinalis
OC   Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
XX
RN   [1]
RP   1-1086
RA   Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Kim U., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
RT   "Giardia: a model for ancient eukaryotic genome analysis";
RL   Unpublished.
XX
RN   [2]
RP   1-1086
RA   Kim U., Morrison H.G., McArthur A.G., Nixon J., Eakin N.Q., Crocker M.C.,
RA   Hinkle G., Holder M.E., Sogin M.L.;
RT   ;
RL   Submitted (13-APR-2000) to the EMBL/GenBank/DDBJ databases.
RL   Josephine Bay Paul Center for Comparative Molecular Biology and Evolution,
RL   Marine Biological Laboratory, 7 MBL Street, Woods Hole, MA 02543-1015, USA
XX
CC   * NOTE: This record contains 1 individual
CC   * sequencing reads that have not been assembled into
CC   * contigs. Runs of N are used to separate the reads
CC   * and the order in which they appear is completely
CC   * arbitrary. Low-pass sequence sampling is useful for
CC   * identifying clones that may be gene-rich and allows
CC   * overlap relationships among clones to be deduced.
CC   * However, it should not be assumed that this clone
CC   * will be sequenced to completion. In the event that
CC   * the record is updated, the accession number will
CC   * be preserved.
CC   *         1     1086: contig of 1086 bp in length.
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .1086
FT                   /db_xref="taxon:5741"
FT                   /organism="Giardia intestinalis"
FT                   /strain="WB-C6"
FT                   /clone="KI0649"
XX
SQ   Sequence 1086 BP; 286 A; 268 C; 287 G; 245 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AC046945/rev   ..

 Align seg 1/1  to reverse of: AC046945  from: 1  to: 1086

       7 ArgLeuThrIleSerAlaLeuLeu 14
         ||||||||||||||||||||||||
     981 CGCTTGACAATCTCGGCTCTTCTG 958
```

```
OM of: US-10-048-197-2 to: EST:*   out_format : pfs

Date: Sep 18, 2002  9:07 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US10048197/runat_17092002_135956_23526/app_query.fasta_1.168
 -DB=EST -QFMT=fastap -SUFFIX=oli.rst -GAPOP=4.500 -GAPEXT=0.050
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
 -DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=40
 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500 -MINLEN=0
 -MAXLEN=2000000000 -USER=US10048197_@CGN1_1_4116 -NCPU=6
 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-10-048-197-2
Query length: 111
Database: EST:*
Database sequences: 13736207
Database length: -1841457050
Search time (sec): 1801.620000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence              Strd Orig    ZScore    EScore Len   ! Documentation  ..
gb_est1:AA563389        +   10.00   169.08   3.59      464  ! AA563389 vl75c02.r1 Knowles Sol
gb_gss:AG148814         +   10.00   165.90   5.40      672  ! AG148814 Pan troglodytes DNA, c
gb_est1:AI049267        +    9.00   150.15   40.70     517  ! AI049267 uc85e06.y1 Sugano mous
gb_est2:BI655178        +    9.00   146.76   62.92     768  ! BI655178 603280984F1 NIH_CGAP_M
gb_gss:BH473529         +    9.00   146.67   63.64     776  ! BH473529 BOGDF02TF BOGD Brassic
gb_est2:BE679654        -    8.00   145.67   72.35     107  ! BE679654 df51e06.y1 Xenopus lae
gb_gss:AZ567757         -    8.00   140.39   142.41    198  ! AZ567757 238PvG04 Pv MBN #26/#2
gb_gss:AZ470169         -    8.00   140.26   144.79    201  ! AZ470169 1M0284B12F Mouse 10kb
gb_est1:AV331944        -    8.00   140.22   145.58    202  ! AV331944 AV331944 RIKEN full-le
gb_est2:BE826421        +    8.00   139.92   151.14    209  ! BE826421 QV4-EN0040-250500-227-
gb_est1:AI231241        +    8.00   139.88   151.94    210  ! AI231241 EST227929 Normalized r
gb_gss:BH530215         -    8.00   139.64   156.72    216  ! BH530215 BOHBI50TF BOHB Brassic
gb_est2:BJ093812        -    8.00   139.56   158.32    218  ! BJ093812 BJ093812 NIBB Mochii r
gb_est1:AA480791        -    8.00   138.63   178.41    243  ! AA480791 ne86d08.s1 NCI_CGAP_Ki
gb_gss:AZ782323         -    8.00   137.45   207.70    279  ! AZ782323 2M0022G01R Mouse 10kb
gb_gss:AZ784705         -    8.00   137.35   210.16    282  ! AZ784705 2M0027G12R Mouse 10kb
gb_est1:AV631270        -    8.00   137.03   219.20    293  ! AV631270 AV631270 Chlamydomonas
gb_est1:AW148390        +    8.00   136.79   225.79    301  ! AW148390 xf11e07.x1 NCI_CGAP_Ki
gb_est1:BB477818        +    8.00   136.79   225.79    301  ! BB477818 BB477818 RIKEN full-le
gb_est2:BJ029356        -    8.00   136.65   229.92    306  ! BJ029356 BJ029356 NIBB Mochii r
gb_est1:AL582559        +    8.00   136.38   238.20    316  ! AL582559 AL582559 LTI_NFL010_BC
gb_est2:BE648930        +    8.00   135.70   259.86    342  ! BE648930 UI-M-BH2.3-aoi-f-07-0-
gb_est2:BG513403        -    8.00   135.65   261.53    344  ! BG513403 dae04c10.x1 NICHD XGC
gb_gss:BH113823         -    8.00   135.50   266.55    350  ! BH113823 RPCI-24-252K17.TV RPCI
gb_est2:BM277543        -    8.00   135.43   269.07    353  ! BM277543 Tm_ad_42B03_SKPL Trich
gb_gss:BH033149         -    8.00   135.31   273.27    358  ! BH033149 RPCI-24-294H20.TV RPCI
gb_gss:BH072663         +    8.00   135.12   279.99    366  ! BH072663 RPCI-24-228C8.TV RPCI-
gb_est1:AW121899        -    8.00   135.07   281.68    368  ! AW121899 UI-M-BH2.3-aoi-f-07-0-
gb_gss:AZ301765         +    8.00   135.00   284.21    371  ! AZ301765 AF-F4F Resistance gene
gb_gss:AZ649818         +    8.00   134.89   288.42    376  ! AZ649818 1M0519H08R Mouse 10kb
gb_est1:BB816077        +    8.00   134.68   296.03    385  ! BB816077 BB816077 RIKEN full-le
gb_est1:AA720873        +    8.00   134.66   296.88    386  ! AA720873 nw12h04.s1 NCI_CGAP_SS
gb_est1:AA872067        -    8.00   134.53   301.96    392  ! AA872067 ob65g05.s1 NCI_CGAP_GC
gb_est2:BJ069087        -    8.00   134.42   306.20    397  ! BJ069087 BJ069087 NIBB Mochii r
gb_gss:AZ046905         +    8.00   134.23   313.85    406  ! AZ046905 nbeb0088N06f CUGI Rice
gb_gss:AQ088615         +    8.00   134.16   316.40    409  ! AQ088615 HS_2208_B1_D02_MR CIT
gb_est2:N39807          +    8.00   134.14   317.25    410  ! N39807 yx93b05.r1 Soares melanc
gb_est2:BJ072077        -    8.00   134.12   318.10    411  ! BJ072077 BJ072077 NIBB Mochii r
gb_gss:AZ132522         +    8.00   133.79   331.76    427  ! AZ132522 OSJNBb0070H11f CUGI
gb_est2:BE685620        -    8.00   133.73   334.32    430  ! BE685620 uu56d10.x1 Soares_t
gb_est1:BB848843        +    8.00   133.65   337.75    434  ! BB848843 BB848843 RIKEN full
gb_gss:B86868           +    8.00   133.58   341.17    438  ! B86868 RPCI11-26H10.TV RPCI-
gb_gss:AQ313383         +    8.00   133.46   346.32    444  ! AQ313383 RPCI11-102O20.TV RP
gb_gss:AZ301770         -    8.00   133.44   347.18    445  ! AZ301770 AF-E8R Resistance g
gb_gss:AQ406823         -    8.00   133.44   347.18    445  ! AQ406823 HS_5092_B1_B04_SP6E

seq_name: gb_est1:AA563389

seq_documentation_block:
LOCUS       AA563389               464 bp    mRNA    linear   EST 18-AUG-1997
DEFINITION  vl75c02.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone IMAGE:978050 5', mRNA sequence.
ACCESSION   AA563389
VERSION     AA563389.1  GI:2334854
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:558778
            Putative full length read
            vector to vector length is 512.
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Mus musculus"
                     /strain="B6D2 F1/J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:978050"
                     /clone_lib="Knowles Solter mouse blastocyst B1"
                     /tissue_type="blastocyst"
                     /dev_stage="embryo (pre-implantation)"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally from mRNA prepared
                     from 800 blastocysts.   Primer: SalI(dT):
                     5'-CGGTCGACCGTCGACCGTTTTTTTTTTTTTTT-3'.  cDNAs  were
                     cloned into the NotI/SalI sites of a pSPORT vector (Life
                     Technologies).  Two different size selections:  B1 (larger
                     inserts) and B3."
BASE COUNT       97 a    109 c    110 g    148 t
ORIGIN


alignment_scores:
            Quality:   10.00              Length:     10
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AA563389   ..

 Align seg 1/1  to: AA563389  from: 1  to: 464

      10 IleSerAlaLeuLeuThrAlaLeuLeuVal 19
```

```
                    ||||||||||||||||||||||||||||
      329 ATCAGTGCTCTGCTAACGGCCCTGCTGGTC 358

seq_name: gb_gss:AG148814

seq_documentation_block:
LOCUS       AG148814                672 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-011C03.TJ, genomic survey
            sequence.
ACCESSION   AG148814
VERSION     AG148814.1  GI:16678492
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-011C03.TJ.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 672)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .672
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-011C03.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT      237 a     87 c     99 g    249 t
ORIGIN


alignment_scores:
            Quality:   10.00             Length:     10
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AG148814    ..

 Align seg 1/1  to: AG148814  from: 1  to: 672

      86 LeuSerLeuIleIleSerPheLeuTrpVal 95
         ||||||||||||||||||||||||||||||
     130 CTTTCATTAATAATATCATTTCTTTGGGTA 159

seq_name: gb_est1:AI049267

seq_documentation_block:
LOCUS       AI049267                517 bp    mRNA    linear   EST 08-JUL-1998
DEFINITION  uc85e06.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:1432450 5', mRNA sequence.
ACCESSION   AI049267
VERSION     AI049267.1  GI:3297554
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:916518
            Seq primer: custom primer used
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1. .517
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1432450"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      108 a    147 c    152 g    110 t
ORIGIN


alignment_scores:
            Quality:    9.00             Length:      9
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AI049267    ..

 Align seg 1/1  to: AI049267  from: 1  to: 517

      13 LeuLeuThrAlaLeuLeuValThrGly 21
         |||||||||||||||||||||||||||
      30 CTGCTTACAGCCCTCCTTGTGACGGGG 56

seq_name: gb_est2:BI655178

seq_documentation_block:
LOCUS       BI655178                768 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603280984F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5325357 5',
            mRNA sequence.
```

```
ACCESSION   BI655178
VERSION     BI655178.1  GI:15569414
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 768)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11825  row: g  column: 22
            High quality sequence stop: 760.
FEATURES             Location/Qualifiers
     source          1. .768
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5325357"
                     /clone_lib="NIH_CGAP_Mam4"
                     /tissue_type="tumor, gross tissue"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert 2.5 kb.  Library constructed by Life
                     Technologies, catalog # 12018-016. Investigators providing
                     samples: Lothar Hennighausen/Priscilla Furth, NIH
                     Reference for transgenic model: Li et al., Cell Growth and
                     Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
                     Library."
BASE COUNT      180 a    214 c    225 g    149 t
ORIGIN


alignment_scores:
           Quality:    9.00            Length:      9
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BI655178   ..

 Align seg 1/1  to: BI655178  from: 1  to: 768

       8 LeuThrIleSerAlaLeuLeuThrAla 16
         ||||||||||||||||||||||||||||
     180 CTCACCATCTCCGCTCTCCTTACAGCA 206

seq_name: gb_gss:BH473529

seq_documentation_block:
LOCUS       BH473529               776 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGDF02TF BOGD Brassica oleracea genomic clone BOGDF02, DNA
            sequence.
ACCESSION   BH473529
VERSION     BH473529.1  GI:17681640
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 776)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGDF02TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .776
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGDF02"
                     /clone_lib="BOGD"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      197 a    142 c    218 g    219 t
ORIGIN


alignment_scores:
           Quality:    9.00            Length:      9
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BH473529   ..

 Align seg 1/1  to: BH473529  from: 1  to: 776

      99 ArgProLysAsnLeuSerPheTyrLeu 107
         |||||||||||||||||||||||||||
     502 AGACCTAAAAATCTTAGTTTTTATCTC 528

seq_name: gb_est2:BE679654

seq_documentation_block:
LOCUS       BE679654               107 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  df51e06.y1 Xenopus laevis unfertilized egg cDNA library Xenopus
            laevis cDNA clone IMAGE:3742882 5', mRNA sequence.
ACCESSION   BE679654
VERSION     BE679654.1  GI:10062553
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 107)
  AUTHORS   Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.
            , Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
            ,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     WashU Xenopus EST project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Sandy Clifton, Ph.D.
            WashU Xenopus EST project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Bruce Blumberg
             Library normalized by Jihwan Song
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: Xenopus clone distribution information for
            this library can be found through Research Genetics, visit their
            web page at: http://www.resgen.com/ Please reference the id listed
```

```
            below when ordering this clone: Source lab clone id - xlnneg001g11
            Seq primer: -40RP from Gibco
            High quality sequence stop: 83.
FEATURES              Location/Qualifiers
     source           1. .107
                      /organism="Xenopus laevis"
                      /db_xref="taxon:8355"
                      /clone="IMAGE:3742882"
                      /clone_lib="Xenopus laevis unfertilized egg cDNA library"
                      /tissue_type="unfertilized egg"
                      /lab_host="Top-10 F'"
                      /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                      XhoI; cDNA was prepared from 2ug of poly A+ RNA.
                      EcoRI-XhoI cut cDNA was then ligated into UniZap-XR
                      (Stratagene) with EcoRI at the 5' end and XhoI at the 3'
                      end. SS-library phagemids were prepared by mass excision
                      from the original library and normalized by  hybridization
                      to biotinylated driver (prepared from the same library by
                      PCR) to Cot-omega of 11. After removal of hybrids and
                      excess driver by streptavidin sepharose chromatography,
                      the  ss-phagemids were made double stranded and
                      electroporated into Top-10 F'. Original library
                      contruction by Bruce Blumberg (Blumberg et al., 1991
                      Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,
                      2923-2935). Note: This is a Xenopus Gene Collection (XGC)
                      library."
BASE COUNT       32 a     17 c     36 g     21 t      1 others
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BE679654/rev   ..

 Align seg 1/1  to reverse of: BE679654  from: 1  to: 107

    11 SerAlaLeuLeuThrAlaLeuLeu 18
       ||||||||||||||||||||||||
    90 TCTGCTCTTCTCACAGCACTGCTG 67

seq_name: gb_gss:AZ567757

seq_documentation_block:
LOCUS       AZ567757                 198 bp    DNA     linear   GSS 07-MAY-2001
DEFINITION  238PvG04 Pv MBN #26/#27 (amplified once) Plasmodium vivax genomic
            3', DNA sequence.
ACCESSION   AZ567757
VERSION     AZ567757.1  GI:13976175
KEYWORDS    GSS.
SOURCE      malaria parasite P. vivax.
  ORGANISM  Plasmodium vivax
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 198)
  AUTHORS   Carlton,J.M.-R. and Dame,J.B.
  TITLE     The Plasmodium vivax and P. berghei gene sequence tag projects
  JOURNAL   Parasitol. Today 16 (10), 409 (2000)
COMMENT     Contact: Dame JB
            Dept. of Pathobiology, College of Veterinary Medicine
            University of Florida
            2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
            Tel: 352 392 4700
            Fax: 352 392 9704
            Email: damej@mail.vetmed.ufl.edu
            Seq primer: M13(-20) forward
            Class: shotgun.
FEATURES              Location/Qualifiers
     source           1. .198
                      /organism="Plasmodium vivax"
                      /strain="Belem"
                      /db_xref="taxon:5855"
                      /clone_lib="Pv MBN #26/#27 (amplified once)"
                      /dev_stage="asexual blood forms"
                      /lab_host="Saimiri boliviensis"
                      /note="Vector: Lambda ZAP II (Stratagene); individual
                      clones excised into phagemid pBluescript; Site_1: EcoR I;
                      Site_2: EcoR I; Genomic DNA was prepared from asynchronous
                      blood stage forms of the Belem line of P. vivax grown in
                      squirrel monkeys. Parasitized erythrocytes were purified
                      from contaminating host leukocytes by filtration of ADP
                      activated blood through acid-washed glass beads and
                      Whatman CF11 cellulose columns by gravity filtration.
                      Purified DNA was digested with mung bean nuclease in the
                      presence of 42.5% formamide at 50oC as described (Galinski
                      , M. et al. 1992. Cell 69,1213-1226; Vernick, K.D. et
                      al.1988. N.A.R. 16, 6883-6896). Eco RI linkers were added
                      and the constructs ligated into Lambda ZAP II. P. vivax
                      Belem was originally isolated from a patient in Belem,
                      Brazil 1980 by Mercia de Arruda, adapted to Saimiri
                      monkeys by Jurg Gysin, and maintained since 1983 in
                      squirrel monkeys."
BASE COUNT       56 a     43 c     62 g     37 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AZ567757/rev   ..

 Align seg 1/1  to reverse of: AZ567757  from: 1  to: 198

    11 SerAlaLeuLeuThrAlaLeuLeu 18
       ||||||||||||||||||||||||
   189 TCTGCGCTTTTGACTGCCCTTCTG 166

seq_name: gb_gss:AZ470169

seq_documentation_block:
LOCUS       AZ470169                 201 bp    DNA     linear   GSS 04-OCT-2000
DEFINITION  1M0284B12F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0284B12 F, DNA sequence.
ACCESSION   AZ470169
VERSION     AZ470169.1  GI:10628294
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 201)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0284  row: B  column: 12
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 201.
```

```
FEATURES             Location/Qualifiers
     source          1. .201
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0284B12"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT       38 a     73 c     50 g     40 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AZ470169/rev   ..

 Align seg 1/1  to reverse of: AZ470169  from: 1  to: 201

      11 SerAlaLeuLeuThrAlaLeuLeu 18
         ||||||||||||||||||||||||
      72 TCTGCACTGCTGACTGCCCTGCTG 49

seq_name: gb_est1:AV331944

seq_documentation_block:
LOCUS        AV331944                202 bp    mRNA    linear   EST 11-NOV-1999
DEFINITION   AV331944 RIKEN full-length enriched, adult male medulla oblongata
             Mus musculus cDNA clone 6330528C24 3', mRNA sequence.
ACCESSION    AV331944
VERSION      AV331944.1  GI:6371996
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 202)
  AUTHORS    Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
             Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
             Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai
             ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
             Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
             Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
             ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
             Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
             Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
             Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE      RIKEN Mouse ESTs (Konno,H., et al. 1999)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Yoshihide Hayashizaki
             Laboratory for Genome Exploration Research Group, RIKEN Genomic
             Sciences Center(GSC), Yokohama Institute
             The Institute of Physical and Chemical Research (RIKEN)
             1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             Tel: 81-45-503-9222
             Fax: 81-45-503-9216
             Email: genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/
             Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
             Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
             ,Y.
              Transcriptional sequencing: A method for DNA sequencing using RNA
             polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
              Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
             Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
             ,Y. and Hayashizaki,Y.
              Automated filtration-based high-throughput plasmid preparation
             system. Genome Res. 9 (5), 463-470 (1999)
              Carninci,P. and Hayashizaki,Y.
              High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
             19-44 (1999)
              Please visit our web site (http://genome.rtc.riken.go.jp) for
             further details.
FEATURES             Location/Qualifiers
     source          1. .202
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="6330528C24"
                     /clone_lib="RIKEN full-length enriched, adult male medulla
                     oblongata"
                     /sex="male"
                     /tissue_type="medulla oblongata"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
BASE COUNT      101 a     20 c     51 g     30 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AV331944/rev   ..

 Align seg 1/1  to reverse of: AV331944  from: 1  to: 202

     103 LeuSerPheTyrLeuThrAlaLys 110
         ||||||||||||||||||||||||
     100 CTTTCTTTCTATCTTACAGCCAAA 77

seq_name: gb_est2:BE826421
```

```
seq_documentation_block:
LOCUS       BE826421                209 bp    mRNA    linear   EST 22-SEP-2000
DEFINITION  QV4-EN0040-250500-227-c11 EN0040 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE826421
VERSION     BE826421.1  GI:10258799
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 209)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV4-EN0040-250
            500-227-c11&t3=2000-05-25&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 87
            High quality sequence stop: 209.
FEATURES             Location/Qualifiers
     source          1. .209
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="EN0040"
                     /dev_stage="Adult"
                     /note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       52 a     64 c     53 g     40 t
ORIGIN


alignment_scores:
           Quality:    8.00            Length:      8
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BE826421   ..

 Align seg 1/1  to: BE826421  from: 1  to: 209

       7 ArgLeuThrIleSerAlaLeuLeu 14
         ||||||||||||||||||||||||
     166 CGGCTTACCATCTCGGCTCTTCTC 189

seq_name: gb_est1:AI231241

seq_documentation_block:
LOCUS       AI231241                210 bp    mRNA    linear   EST 31-JAN-1999
DEFINITION  EST227929 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
            REMDH53 3' end, mRNA sequence.
ACCESSION   AI231241
VERSION     AI231241.1  GI:3815121
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 210)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1. .210
                     /organism="Rattus sp."
                     /db_xref="ATCC (inhost):2037375"
                     /db_xref="taxon:10118"
                     /clone="REMDH53"
                     /clone_lib="Normalized rat embryo, Bento Soares"
                     /dev_stage="embryo 8, 12, 18 dpc"
                     /note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT       51 a     42 c     37 g     80 t
ORIGIN


alignment_scores:
           Quality:    8.00            Length:      8
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AI231241   ..

 Align seg 1/1  to: AI231241  from: 1  to: 210

      85 IleLeuSerLeuIleIleSerPhe 92
         |||||||||||||||||||||||
      85 ATTCTTTCCTTGATAATCTCTTTT 108

seq_name: gb_gss:BH530215

seq_documentation_block:
LOCUS       BH530215                216 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOHBI50TF BOHB Brassica oleracea genomic clone BOHBI50, DNA
            sequence.
ACCESSION   BH530215
VERSION     BH530215.1  GI:17746359
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 216)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOHBI50TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
```

```
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .216
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHBI50"
                     /clone_lib="BOHB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       69 a     18 c     50 g     79 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BH530215/rev   ..

 Align seg 1/1  to reverse of: BH530215  from: 1  to: 216

      86 LeuSerLeuIleIleSerPheLeu 93
         |||||||||||||||||||||||||
     181 CTAAGTCTAATTATTTCCTTTCTA 158

seq_name: gb_est2:BJ093812

seq_documentation_block:
LOCUS        BJ093812                218 bp    mRNA    linear   EST 12-DEC-2001
DEFINITION   BJ093812 NIBB Mochii normalized Xenopus early gastrula library
             Xenopus laevis cDNA clone XL143j12 5', mRNA sequence.
ACCESSION    BJ093812
VERSION      BJ093812.1  GI:17593552
KEYWORDS     EST.
SOURCE       African clawed frog.
  ORGANISM   Xenopus laevis
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
             Xenopodinae; Xenopus.
REFERENCE    1  (bases 1 to 218)
  AUTHORS    Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
             ,Y.
  TITLE      Expressed genes in X. laevis embryo
  JOURNAL    Unpublished (2001)
COMMENT      Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .218
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL143j12"
                     /clone_lib="NIBB Mochii normalized Xenopus early gastrula
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 10.5"
BASE COUNT       56 a     46 c     58 g     56 t      2 others
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BJ093812/rev   ..

 Align seg 1/1  to reverse of: BJ093812  from: 1  to: 218

      11 SerAlaLeuLeuThrAlaLeuLeu 18
         ||||||||||||||||||||||||
      65 TCTGCTCTTCTCACAGCACTGCTG 42

seq_name: gb_est1:AA480791

seq_documentation_block:
LOCUS        AA480791                243 bp    mRNA    linear   EST 14-AUG-1997
DEFINITION   ne86d08.s1 NCI_CGAP_Kid1 Homo sapiens cDNA clone IMAGE:911151, mRNA
             sequence.
ACCESSION    AA480791
VERSION      AA480791.1  GI:2209489
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 243)
  AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
             Emmert-Buck, M.D., Ph.D.
              cDNA Library Preparation: David B. Krizman, Ph.D.
              cDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Insert Length: 427   Std Error: 0.00
             Seq primer: -41m13 fwd. ET from Amersham.
FEATURES             Location/Qualifiers
     source          1. .243
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:911151"
                     /clone_lib="NCI_CGAP_Kid1"
                     /tissue_type="kidney"
                     /lab_host="DH10B"
                     /note="Vector: pAMP10; mRNA made from invasive kidney
                     tumor, cDNA made by oligo-dT priming.  Non-directionally
                     cloned.  Size-selected on agarose gel, average insert
                     size 600 bp. Reference:  Krizman et al. (1996) Cancer
                     Research 56:5380-5383."
BASE COUNT       98 a     56 c     50 g     39 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AA480791/rev   ..

 Align seg 1/1  to reverse of: AA480791  from: 1  to: 243

      52 IleSerSerArgPheGlySerAla 59
         |||||||||||||||||||||||||
     131 ATTTCTTCTCGGTTTGGGTCTGCT 108

seq_name: gb_gss:AZ782323
```

```
seq_documentation_block:
LOCUS       AZ782323                279 bp    DNA     linear   GSS 16-FEB-2001
DEFINITION  2M0022G01R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0022G01 R, DNA sequence.
ACCESSION   AZ782323
VERSION     AZ782323.1  GI:12915930
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 279)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0022  row: G  column: 01
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 279.
FEATURES             Location/Qualifiers
     source          1. .279
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC2M0022G01"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      128 a     46 c     52 g     53 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AZ782323/rev   ..

 Align seg 1/1  to reverse of: AZ782323  from: 1  to: 279

      86 LeuSerLeuIleIleSerPheLeu 93
         ||||||||||||||||||||||||
     217 CTGAGTTTGATTATTTCCTTCCTT 194

 seq_name: gb_gss:AZ784705

 seq_documentation_block:
 LOCUS       AZ784705                282 bp    DNA     linear   GSS 16-FEB-2001
 DEFINITION  2M0027G12R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
             clone UUGC2M0027G12 R, DNA sequence.
 ACCESSION   AZ784705
 VERSION     AZ784705.1  GI:12920711
 KEYWORDS    GSS.
 SOURCE      house mouse.
   ORGANISM  Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE   1  (bases 1 to 282)
   AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
             Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
             ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
             and Wright,D.,Weiss,R.
   TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
             plasmid inserts
   JOURNAL   Unpublished (2000)
 COMMENT     Contact: Robert B. Weiss
             University of Utah Genome Center
             University of Utah
             Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
             84112, USA
             Tel: 801 585 5606
             Fax: 801 585 7177
             Email: ddunn@genetics.utah.edu
             Insert Length: 10000   Std Error: 0.00
             Plate: 0027  row: G  column: 12
             Seq primer: CACACAGGAAACAGCTATGACC
             Class: plasmid ends
             High quality sequence stop: 282.
 FEATURES             Location/Qualifiers
      source          1. .282
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC2M0027G12"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
 BASE COUNT       66 a     94 c     80 g     42 t
 ORIGIN


 alignment_scores:
             Quality:    8.00            Length:      8
```

```
            Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AZ784705/rev   ..

 Align seg 1/1  to reverse of: AZ784705  from: 1  to: 282

      11 SerAlaLeuLeuThrAlaLeuLeu 18
         ||||||||||||||||||||||||
     228 TCTGCACTGCTGACTGCCCTGCTG 205

seq_name: gb_est1:AV631270

seq_documentation_block:
LOCUS       AV631270                293 bp    mRNA    linear   EST 15-DEC-2000
DEFINITION  AV631270 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
            reinhardtii cDNA clone LCL091b11_r 5', mRNA sequence.
ACCESSION   AV631270
VERSION     AV631270.1  GI:10793904
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 293)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .293
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="LCL091b11_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a carbon stress acclimatized condition in which carbon
                     dioxide concentration in the bubbling gas was changed from
                     5% to 0.04%"
BASE COUNT       70 a     85 c     89 g     49 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AV631270/rev   ..

 Align seg 1/1  to reverse of: AV631270  from: 1  to: 293

       7 ArgLeuThrIleSerAlaLeuLeu 14
         ||||||||||||||||||||||||
     207 CGCTTAACGATTTCTGCATTGTTG 184

seq_name: gb_est1:AW148390

seq_documentation_block:
LOCUS       AW148390                301 bp    mRNA    linear   EST 03-NOV-1999
DEFINITION  xf11e07.x1 NCI_CGAP_Kid8 Homo sapiens cDNA clone IMAGE:2617764 3'
            similar to contains Alu repetitive element;contains element MSR1
            MSR1 repetitive element ;, mRNA sequence.
ACCESSION   AW148390
VERSION     AW148390.1  GI:6196211
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 301)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40UP from Gibco
            High quality sequence stop: 301.
FEATURES             Location/Qualifiers
     source          1. .301
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2617764"
                     /clone_lib="NCI_CGAP_Kid8"
                     /tissue_type="renal cell tumor"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.2 kb. Life Technologies catalog #:
                     11524-014"
BASE COUNT       53 a     90 c     76 g     80 t      2 others
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AW148390   ..

 Align seg 1/1  to: AW148390  from: 1  to: 301

       9 ThrIleSerAlaLeuLeuThrAla 16
         ||||||||||||||||||||||||
      55 ACGATCTCTGCTTTGCTTACTGCA 78

seq_name: gb_est1:BB477818

seq_documentation_block:
LOCUS       BB477818                301 bp    mRNA    linear   EST 23-JUL-2000
DEFINITION  BB477818 RIKEN full-length enriched, 13 days embryo heart Mus
            musculus cDNA clone D330027A14 3', mRNA sequence.
ACCESSION   BB477818
VERSION     BB477818.1  GI:9395427
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 301)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
            ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
```

```
                Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.
                , Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
                , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
                ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.
                , Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tominaga,N., Toya
                ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.
                , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
                ,M., Muramatsu,M. and Hayashizaki,Y.
  TITLE         RIKEN Mouse ESTs (Konno,H., et al.)
  JOURNAL       Unpublished (2000)
  COMMENT       Contact: Yoshihide Hayashizaki
                Laboratory for Genome Exploration Research Group, RIKEN Genomic
                Sciences Center(GSC), Yokohama Institute
                The Institute of Physical and Chemical Research (RIKEN)
                1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
                Tel: 81-45-503-9222
                Fax: 81-45-503-9216
                Email: genome-res@gsc.riken.go.jp,
                URL:http://genome.gsc.riken.go.jp/
                Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
                ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
                Thermostabilization and thermoactivation of thermolabile enzymes by
                trehalose and its application for the synthesis of full length
                cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
                Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
                Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
                ,Y. and Hayashizaki,Y.
                Automated filtration-based high-throughput plasmid preparation
                system. Genome Res. 9 (5), 463-470 (1999)
                Carninci,P. and Hayashizaki,Y.
                High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
                19-44 (1999)
                Please visit our web site (http://genome.rtc.riken.go.jp) for
                further details.
  FEATURES              Location/Qualifiers
       source           1. .301
                        /organism="Mus musculus"
                        /db_xref="taxon:10090"
                        /clone="D330027A14"
                        /clone_lib="RIKEN full-length enriched, 13 days embryo
                        heart"
                        /tissue_type="heart"
                        /dev_stage="13 days embryo"
                        /lab_host="DH10B"
                        /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                        prepared and sequenced in Mouse Genome Encyclopedia
                        Project of Genome Exploration Research Group in Riken
                        Genomic Sciences Center and Genome Science Laboratory in
                        RIKEN. Division of Experimental Animal Research in Riken
                        contributed to prepare mouse tissues. 1st strand cDNA was
                        primed with a primer [5'
                        GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
                        prepared by using trehalose thermo-activated reverse
                        transcriptase and subsequently enriched for full-length by
                        cap-trapper. Second strand cDNA was prepared with the
                        primer adapter of sequence [5'
                        GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA
                        was cleaved with BamHI and XhoI. Vector: a modified
                        pBluescript KS(+) after bulk excision from Lambda FLC I."
  BASE COUNT        88 a     59 c     59 g     95 t
  ORIGIN


  alignment_scores:
              Quality:    8.00              Length:      8
                Ratio:  1.000                Gaps:      0
   Percent Similarity: 100.000   Percent Identity: 100.000

  alignment_block:
   US-10-048-197-2 x BB477818   ..

   Align seg 1/1  to: BB477818  from: 1  to: 301
```

```
       22 CysValSerThrGlyAsnValAla 29
          ||||||||||||||||||||||||
       73 TGTGTCAGTACTGGAAATGTGGCC 96

  seq_name: gb_est2:BJ029356

  seq_documentation_block:
  LOCUS        BJ029356                306 bp    mRNA    linear   EST 07-DEC-2001
  DEFINITION   BJ029356 NIBB Mochii normalized Xenopus neurula library Xenopus
               laevis cDNA clone XL012d07 5', mRNA sequence.
  ACCESSION    BJ029356
  VERSION      BJ029356.1  GI:17418919
  KEYWORDS     EST.
  SOURCE       African clawed frog.
    ORGANISM   Xenopus laevis
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
               Xenopodinae; Xenopus.
  REFERENCE    1  (bases 1 to 306)
    AUTHORS    Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
               ,Y.
    TITLE      Expressed genes in X. laevis embryo
    JOURNAL    Unpublished (2001)
  COMMENT      Contact: Tadasu Shin-i
               Center For Genetic Resource Information
               National Institute of Genetics
               1111 Yata, Mishima, Shizuoka 411-8540, Japan
               Tel: 81-559-81-6856
               Fax: 81-559-81-6855
               Email: tshini@genes.nig.ac.jp.
  FEATURES             Location/Qualifiers
       source          1. .306
                       /organism="Xenopus laevis"
                       /db_xref="taxon:8355"
                       /clone="XL012d07"
                       /clone_lib="NIBB Mochii normalized Xenopus neurula
                       library"
                       /tissue_type="whole embryo"
                       /dev_stage="stage 15"
  BASE COUNT       87 a     56 c     83 g     80 t
  ORIGIN


  alignment_scores:
              Quality:    8.00              Length:      8
                Ratio:  1.000                Gaps:      0
   Percent Similarity: 100.000   Percent Identity: 100.000

  alignment_block:
   US-10-048-197-2 x BJ029356/rev   ..

   Align seg 1/1  to reverse of: BJ029356  from: 1  to: 306

       11 SerAlaLeuLeuThrAlaLeuLeu 18
          ||||||||||||||||||||||||
      151 TCTGCTCTTCTCACAGCACTGCTG 128

  seq_name: gb_est1:AL582559

  seq_documentation_block:
  LOCUS        AL582559                316 bp    mRNA    linear   EST 16-FEB-2001
  DEFINITION   AL582559 LTI_NFL010_BC2 Homo sapiens cDNA clone CS0DL010YK13 5
               prime, mRNA sequence.
  ACCESSION    AL582559
  VERSION      AL582559.1  GI:12950661
  KEYWORDS     EST.
  SOURCE       human.
    ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE    1  (bases 1 to 316)
    AUTHORS    Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
    TITLE      Full-length cDNA libraries and normalization
```

```
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .316
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DL010YK13"
                     /clone_lib="LTI_NFL010_BC2"
                     /sex="male"
                     /tissue_type="B cells from Burkitt lymphoma"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT       63 a    101 c     77 g     68 t      7 others
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AL582559   ..

 Align seg 1/1  to: AL582559  from: 1  to: 316

      12 AlaLeuLeuThrAlaLeuLeuVal 19
         ||||||||||||||||||||||||
     193 GCCCTGCTAACAGCACTGCTGGTA 216

seq_name: gb_est2:BE648930

seq_documentation_block:
LOCUS       BE648930               342 bp    mRNA    linear   EST 06-SEP-2000
DEFINITION  UI-M-BH2.3-aoi-f-07-0-UI.r1 NIH_BMAP_M_S3.3 Mus musculus cDNA clone
            UI-M-BH2.3-aoi-f-07-0-UI 5', mRNA sequence.
ACCESSION   BE648930
VERSION     BE648930.1  GI:9974751
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 342)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .342
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH2.3-aoi-f-07-0-UI"
                     /clone_lib="NIH_BMAP_M_S3.3"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S3.3 library is a subtracted library of a
                     series,  ultimately derived from a mixture of individually
                     tagged normalized  libraries from ten regions of the mouse
                     brain (cerebellum, brain  stems, olfactory bulbs,
                     hypothalamus, cortex, amygdala, basal  ganglia, pineal
                     gland, striatum, hipoccampus) after a series of
                     subtractions to reduce the representation of cDNAs from
                     which ESTs  had already been generated.  The following
                     serially subtracted  libraries were generated in this
                     process: NIH_BMAP_M_S3.3,  NIH_BMAP_M_S2, NIH_BMAP_M_S1.
                     The subtracted library (NIH_BMAP_M_S3.3)  was constructed
                     as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S2
                     clones from which 3' ESTs had been derived was used as  a
                     driver in a hybridization with the NIH_BMAP_M_S2 library
                     in the form  of single-stranded circles.  The remaining
                     single-stranded circles  (subtracted library) was purified
                     by hydroxyapatite column  chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (LifeTechnologies) to generate the
                     NIH_BMAP_M_S3.3 library. This procedure has been
                     previously described (Bonaldo,  Lennon and Soares, Genome
                     Research 6:791-806, 1996)"
BASE COUNT       72 a     71 c     76 g    123 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BE648930   ..

 Align seg 1/1  to: BE648930  from: 1  to: 342

      22 CysValSerThrGlyAsnValAla 29
         ||||||||||||||||||||||||
     315 TGTGTCAGTACTGGAAATGTGGCA 338

seq_name: gb_est2:BG513403

seq_documentation_block:
LOCUS       BG513403               344 bp    mRNA    linear   EST 28-MAR-2001
DEFINITION  dae04c10.x1 NICHD XGC Lul Xenopus laevis cDNA clone IMAGE:4632979
            3', mRNA sequence.
ACCESSION   BG513403
VERSION     BG513403.1  GI:13484060
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 344)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

```
            Tissue Procurement: Martha Rebbert, Steven L. Klein, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: Xenopus clones from this library are available
            through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .344
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:4632979"
                     /clone_lib="NICHD XGC Lu1"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.6 kb. Constructed by Life
                     Technologies. Note: This is a Xenopus Gene Collection (XGC
                     ) library."
BASE COUNT      118 a     70 c     84 g     72 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BG513403/rev   ..

 Align seg 1/1  to reverse of: BG513403  from: 1  to: 344

      10 IleSerAlaLeuLeuThrAlaLeu 17
         ||||||||||||||||||||||||
     106 ATCAGTGCACTCCTCACTGCACTG 83

seq_name: gb_gss:BH113823

seq_documentation_block:
LOCUS       BH113823               350 bp    DNA     linear   GSS 19-JUL-2001
DEFINITION  RPCI-24-252K17.TV RPCI-24 Mus musculus genomic clone RPCI-24-252K17
            , DNA sequence.
ACCESSION   BH113823
VERSION     BH113823.1  GI:14951386
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 350)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-252K17.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 252  row: K  column: 17
            Seq primer: T7
            Class: BAC ends.

FEATURES             Location/Qualifiers
     source          1. .350
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-252K17"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT      142 a     53 c     69 g     86 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BH113823/rev   ..

 Align seg 1/1  to reverse of: BH113823  from: 1  to: 350

      86 LeuSerLeuIleIleSerPheLeu 93
         |||||||||||||||||||||||||
     297 TTATCTCTGATTATTTCCTTTCTG 274

seq_name: gb_est2:BM277543

seq_documentation_block:
LOCUS       BM277543               353 bp    mRNA    linear   EST 20-DEC-2001
DEFINITION  Tm_ad_42B03_SKPL Trichuris muris (parasitic nematode) mixed adult
            Trichuris muris cDNA clone Tm_ad_42B03 5' similar to gb|AAC02985.2|
            (AF038554) density regulated protein drp1 - Homo sapiens, mRNA
            sequence.
ACCESSION   BM277543
VERSION     BM277543.1  GI:17970794
KEYWORDS    EST.
SOURCE      Trichuris muris.
  ORGANISM  Trichuris muris
            Eukaryota; Metazoa; Nematoda; Enoplea; Trichocephalida; Trichuridae
            ; Trichuris.
REFERENCE   1  (bases 1 to 353)
  AUTHORS   Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Guiliano,D., Hall
            ,N., Quayle,M. and Barrell,B.
  TITLE     Edinburgh University/Sanger Centre Nematode EST Project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Blaxter ML
            Institute of Cell, Animal and Population Biology
            University of Edinburgh
            Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
            3JT, UK.
            Tel: +44 131 650 6760
            Fax: +44 131 670 5450
            Email: mark.blaxter@ed.ac.uk
            The library was prepared by Richard Grencis, Manchester University,
            Manchester. Sequencing was performed by the Pathogen Sequencing
            Unit, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart
            Barrell).
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7PL
            Plate: 42  row: B  column: 03
            Seq primer: SKPL
            High quality sequence stop: 278.
FEATURES             Location/Qualifiers
     source          1. .353
                     /organism="Trichuris muris"
```

```
                     /db_xref="taxon:70415"
                     /clone="Tm_ad_42B03"
                     /clone_lib="Trichuris muris (parasitic nematode) mixed
                     adult"
                     /sex="mixed"
                     /dev_stage="adult"
                     /note="Vector: Lambda Zap II; Site_1: EcoRI (5'end);
                     Site_2: XhoI (3'end); Trichuris muris is a nematode
                     parasite of rodents related to the human whipworm
                     Trichuris trichiura. The library was constructed from
                     Trichuris muris adults (Edinburgh 'E' strain) maintained
                     in mice, and was provided by Dr. Richard Grencis,
                     University of Manchester."
BASE COUNT      102 a     80 c     99 g     62 t     10 others
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BM277543/rev   ..

 Align seg 1/1  to reverse of: BM277543  from: 1  to: 353

      19 ValThrGlyCysValSerThrGly 26
         ||||||||||||||||||||||||
     249 GTAACAGGCTGTGTTTCGACTGGC 226

seq_name: gb_gss:BH033149

seq_documentation_block:
LOCUS       BH033149                358 bp    DNA     linear   GSS 17-JUL-2001
DEFINITION  RPCI-24-294H20.TV RPCI-24 Mus musculus genomic clone RPCI-24-294H20
            , DNA sequence.
ACCESSION   BH033149
VERSION     BH033149.1  GI:14804691
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 358)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 294  row: H  column: 20
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .358
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-294H20"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
BASE COUNT       99 a     60 c     76 g    123 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BH033149/rev   ..

 Align seg 1/1  to reverse of: BH033149  from: 1  to: 358

      86 LeuSerLeuIleIleSerPheLeu 93
         |||||||||||||||||||||||||
     307 CTGTCACTCATAATTTCATTCCTT 284

seq_name: gb_gss:BH072663

seq_documentation_block:
LOCUS       BH072663                366 bp    DNA     linear   GSS 18-JUL-2001
DEFINITION  RPCI-24-228C8.TV RPCI-24 Mus musculus genomic clone RPCI-24-228C8,
            DNA sequence.
ACCESSION   BH072663
VERSION     BH072663.1  GI:14892260
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 366)
  AUTHORS   Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
            Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
            Russell,D., de Jong,P. and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-24
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-24-228C8.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-24. For BAC
            library availability, please contact Pieter de Jong
            (pdejong@mail.cho.org). Clones may be purchased from BACPAC
            Resources (http://www.chori.org/bacpac/orderingframe.htm). BAC end
            page: http://ww .tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 228  row: C  column: 8
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .366
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-24-228C8"
                     /clone_lib="RPCI-24"
                     /sex="Male"
                     /cell_type="Spleen/Brain"
                     /note="Vector: pTARBAC1; Site_1: BamH1; Site_2: BamH1;
                     RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
                     library was  cloned in the pTARBAC1 cloning vector at the
                     BamH1 sites using MboI  partially digested male C57BL/6J
                     DNA."
```

```
BASE COUNT       56 a     76 c     68 g    166 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BH072663   ..

 Align seg 1/1  to: BH072663  from: 1  to: 366

      86 LeuSerLeuIleIleSerPheLeu 93
         |||||||||||||||||||||||||
     166 CTGAGTTTGATTATTTCCTTTCTC 189

seq_name: gb_est1:AW121899

seq_documentation_block:
LOCUS       AW121899                368 bp    mRNA    linear   EST 22-OCT-1999
DEFINITION  UI-M-BH2.3-aoi-f-07-0-UI.s1 NIH_BMAP_M_S3.3 Mus musculus cDNA clone
            UI-M-BH2.3-aoi-f-07-0-UI 3', mRNA sequence.
ACCESSION   AW121899
VERSION     AW121899.1  GI:6097273
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized brain stems library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
            available by the means that is soon to be determined. When NIH
            determines the means for distribution of the BMAP cDNA clones, this
            record will be updated accordingly when that means is determined.
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH2.3-aoi-f-07-0-UI"
                     /clone_lib="NIH_BMAP_M_S3.3"
                     /dev_stage="27-32 days"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NIH_BMAP_M_S3.3 library is a subtracted library of a
                     series,  ultimately derived from a mixture of individually
                     tagged normalized  libraries from ten regions of the mouse
                     brain (cerebellum, brain  stems, olfactory bulbs,
                     hypothalamus, cortex, amygdala, basal  ganglia, pineal
                     gland, striatum, hipoccampus) after a series of
                     subtractions to reduce the representation of cDNAs from
                     which ESTs  had already been generated.  The following
                     serially subtracted  libraries were generated in this
                     process: NIH_BMAP_M_S3.3,  NIH_BMAP_M_S2, NIH_BMAP_M_S1.
                     The subtracted library (NIH_BMAP_M_S3.3)  was constructed
                     as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S2
                     clones from which 3' ESTs had been derived was used as  a
                     driver in a hybridization with the NIH_BMAP_M_S2 library
                     in the form  of single-stranded circles.  The remaining
                     single-stranded circles  (subtracted library) was purified
                     by hydroxyapatite column  chromatography, converted to
                     double-stranded circles and electroporated into DH10B
                     bacteria (LifeTechnologies) to generate the
                     NIH_BMAP_M_S3.3 library. This procedure has been
                     previously described (Bonaldo,  Lennon and Soares, Genome
                     Research 6:791-806, 1996)
                     TAG_LIB=NIH_BMAP_M_S3.3
                     TAG_TISSUE=brain-stems
                     TAG_SEQ=TCATT"
BASE COUNT      119 a     79 c     63 g    107 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AW121899/rev   ..

 Align seg 1/1  to reverse of: AW121899  from: 1  to: 368

      22 CysValSerThrGlyAsnValAla 29
         |||||||||||||||||||||||||
     243 TGTGTCAGTACTGGAAATGTGGCC 220

seq_name: gb_gss:AZ301765

seq_documentation_block:
LOCUS       AZ301765                371 bp    DNA     linear   GSS 23-AUG-2000
DEFINITION  AF-F4F Resistance gene analog sequences of soybean (nonTIR-NBS-LRR
            type) similar to AF060192. Glycine max genomic, DNA sequence.
ACCESSION   AZ301765
VERSION     AZ301765.1  GI:9886539
KEYWORDS    GSS.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   Penuela,S. and Young,N.D.
  TITLE     Resistance gene analog sequences of soybean, Glycine max
            (nonTIR-NBS-LRR type)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Young Nevin D
            Department of Plant Pathology
            University of Minnesota
            495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul MN 55108, USA
            Tel: 612 625 2225
            Fax: 612 625 9728
            Email: neviny@tc.umn.edu
            nonTIR-NBS-LRR RGA sequence of soybean on linkage group F.
            Seq primer: M13F
            Class: RFLP probe.
FEATURES             Location/Qualifiers
     source          1. .371
                     /organism="Glycine max"
                     /cultivar="Faribault"
                     /db_xref="taxon:3847"
                     /clone_lib="Resistance gene analog sequences of soybean
```

(nonTIR-NBS-LRR type) similar to AF060192."
/tissue_type="cotyledon leaves"
/dev_stage="cotyledon stage"
/note="Vector: pCR2.1-TOPO; Site_1: EcoRI; Soybean genomic DNA (cv. Faribault) was amplified by PCR with primers corresponding to the 'P-loop' and the region preceding the 'GLPLA' motif of the nucleotide-binding site (NBS) of soybean GenBank accession AF060192. PCR products of approximately 450 bp were cloned into pCR2.1-TOPO vector, and sequenced with M13 universal primers."
BASE COUNT 106 a 61 c 92 g 98 t 14 others
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AZ301765 ..

Align seg 1/1 to: AZ301765 from: 1 to: 371

50 GlnGluIleSerSerArgPheGly 57
||||||||||||||||||||||||
224 CAAGAGATTTCTTCTCGTTTTGGA 247

seq_name: gb_gss:AZ649818

seq_documentation_block:
LOCUS AZ649818 376 bp DNA linear GSS 14-DEC-2000
DEFINITION 1M0519H08R Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0519H08 R, DNA sequence.
ACCESSION AZ649818
VERSION AZ649818.1 GI:11783680
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 376)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0519 row: H column: 08
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 376.
FEATURES Location/Qualifiers
source 1. .376
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0519H08"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 89 a 108 c 46 g 133 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AZ649818 ..

Align seg 1/1 to: AZ649818 from: 1 to: 376

86 LeuSerLeuIleIleSerPheLeu 93
|||||||||||||||||||||||||
131 CTCTCTCTGATAATTTCTTTTTTA 154

seq_name: gb_est1:BB816077

seq_documentation_block:
LOCUS BB816077 385 bp mRNA linear EST 19-NOV-2001
DEFINITION BB816077 RIKEN full-length enriched, lung RCB-0558 LLC cDNA Mus musculus cDNA clone G730032I23 3', mRNA sequence.
ACCESSION BB816077
VERSION BB816077.1 GI:16988706
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 385)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara ,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

```
FEATURES             Location/Qualifiers
     source          1. .385
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="G730032I23"
                     /clone_lib="RIKEN full-length enriched, lung RCB-0558 LLC
                     cDNA"
                     /tissue_type="lung"
                     /cell_line="RCB-0558 LLC"
BASE COUNT      102 a     69 c     86 g    128 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BB816077   ..

 Align seg 1/1  to: BB816077  from: 1  to: 385

      22 CysValSerThrGlyAsnValAla 29
         ||||||||||||||||||||||||
     152 TGTGTCAGTACTGGAAATGTGGCC 175

seq_name: gb_est1:AA720873

seq_documentation_block:
LOCUS        AA720873                386 bp    mRNA    linear   EST 18-FEB-1998
DEFINITION   nw12h04.s1 NCI_CGAP_SS1 Homo sapiens cDNA clone IMAGE:1239127 3',
             mRNA sequence.
ACCESSION    AA720873
VERSION      AA720873.1  GI:2737008
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 386)
  AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
             Emmert-Buck, M.D., Ph.D.
              cDNA Library Preparation: Stratagene, Inc.
              cDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Insert Length: 2417    Std Error: 0.00
             Seq primer: -40m13 fwd. ET from Amersham
             High quality sequence stop: 377.
FEATURES             Location/Qualifiers
     source          1. .386
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1239127"
                     /clone_lib="NCI_CGAP_SS1"
                     /tissue_type="synovial sarcoma"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI
                     ; Cloned unidirectionally.  Primer: Oligo dT. Synovial
                     sarcoma. 5' adaptor sequence: 5' GAATTCGGCACGAG 3' 3'
                     adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3' Average
                     insert size:  1.4 kb."
BASE COUNT      73 a     62 c     72 g    179 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AA720873   ..

 Align seg 1/1  to: AA720873  from: 1  to: 386

      51 GluIleSerSerArgPheGlySer 58
         |||||||||||||||||||||||||
     317 GAGATTTCTTCTAGGTTTGGATCC 340

seq_name: gb_est1:AA872067

seq_documentation_block:
LOCUS        AA872067                392 bp    mRNA    linear   EST 25-MAR-1998
DEFINITION   ob65g05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1336280 3'
             similar to gb:M55053 CYTOCHROME P450 IA2 (HUMAN);contains Alu
             repetitive element;contains element L1 repetitive element ;, mRNA
             sequence.
ACCESSION    AA872067
VERSION      AA872067.1  GI:2968245
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 392)
  AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
             Ph.D., Gerald Marti, M.D.
              cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
             Bonaldo, Ph.D.
              cDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Insert Length: 1594    Std Error: 0.00
             Seq primer: -40m13 fwd. ET from Amersham
             High quality sequence stop: 304.
FEATURES             Location/Qualifiers
     source          1. .392
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1336280"
                     /clone_lib="NCI_CGAP_GCB1"
                     /tissue_type="germinal center B cell"
```

```
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from human tonsillar cells enriched for
                     germinal center B cells by flow sorting (CD20+, IgD-),
                     provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
                     (NCI) and Dr. Gerald Marti (CBER).  cDNA synthesis was
                     primed with a Not I - oligo(dT) primer
                     [5'-TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCATTTTTTTTTTTTTTTTTT-3'
                     ]. Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      118 a    105 c    109 g     60 t
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AA872067/rev   ..

 Align seg 1/1  to reverse of: AA872067  from: 1  to: 392

       9 ThrIleSerAlaLeuLeuThrAla 16
         ||||||||||||||||||||||||
     347 ACAATCTCTGCTCTGCTCACTGCA 324

seq_name: gb_est2:BJ069087

seq_documentation_block:
LOCUS       BJ069087                397 bp    mRNA    linear   EST 11-DEC-2001
DEFINITION  BJ069087 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL051116 5', mRNA sequence.
ACCESSION   BJ069087
VERSION     BJ069087.1  GI:17494784
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 397)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .397
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL051116"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
BASE COUNT      118 a     81 c    105 g     92 t      1 others
ORIGIN


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BJ069087/rev   ..

 Align seg 1/1  to reverse of: BJ069087  from: 1  to: 397

      11 SerAlaLeuLeuThrAlaLeuLeu 18
         ||||||||||||||||||||||||
     245 TCTGCTCTTCTCACAGCACTGCTG 222

seq_name: gb_gss:AZ046905

seq_documentation_block:
LOCUS       AZ046905                406 bp    DNA     linear   GSS 16-MAR-2000
DEFINITION  nbeb0088N06f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone OSJNBb0088N06f, DNA sequence.
ACCESSION   AZ046905
VERSION     AZ046905.1  GI:7250674
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634., USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GTAAAACGACGGCCAGTG
            Class: BAC ends
            High quality sequence start: 20
            High quality sequence stop: 353.
FEATURES             Location/Qualifiers
     source          1. .406
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="OSJNBb0088N06f"
                     /clone_lib="CUGI Rice BAC Library (EcoRI)"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                     Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Monocotyledonous rice is a diploid plant (2n=24) with a
                     haploid genome equivalent of 431 Mbp (Arumuganathan and
                     Earle, 1991). The relatively small genome of rice, three
                     times larger than that of Arabidopsis, makes it suitable
                     for genomic studies. In order to facilitate positional
                     cloning, physical mapping and genome sequencing of rice,
                     we have constructed a BAC library from Oryza sativa,
                     Nipponbare variety using EcoRI as the cloning enzyme. The
                     library contains 55,296 clones with an average insert size
                     of 121 Kb providing approximatley 15 haploid genome
                     equivalents. The deep coverage allows the isolation a
                     particular sequence with a probability of 99.9 %. Three
                     high density filters, each containing 18,432 clones
                     (doubly spotted), represent the whole library for colony
                     screening and can be requested from the Clemson University
                     BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT      111 a     72 c    130 g     93 t
```

```
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AZ046905   ..

 Align seg 1/1  to: AZ046905  from: 1  to: 406

      86 LeuSerLeuIleIleSerPheLeu 93
         |||||||||||||||||||||||||
      59 TTGAGTTTAATAATTTCCTTTTTG 82

seq_name: gb_gss:AQ088615

seq_documentation_block:
LOCUS       AQ088615                409 bp    DNA     linear   GSS 26-AUG-1998
DEFINITION  HS_2208_B1_D02_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2208 Col=3 Row=H, DNA sequence.
ACCESSION   AQ088615
VERSION     AQ088615.1  GI:3457526
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 409)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 2208  row: H  column: 3
            Class: BAC ends
            High quality sequence stop: 409.
FEATURES             Location/Qualifiers
     source          1. .409
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=2208 Col=3 Row=H"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT       94 a    107 c     93 g    114 t      1 others
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AQ088615   ..

 Align seg 1/1  to: AQ088615  from: 1  to: 409

       7 ArgLeuThrIleSerAlaLeuLeu 14
         ||||||||||||||||||||||||
      76 AGGTTAACTATCTCTGCTTTGCTG 99

seq_name: gb_est2:N39807

seq_documentation_block:
LOCUS       N39807                  410 bp    mRNA    linear   EST 22-JAN-1996
DEFINITION  yx93b05.r1 Soares melanocyte 2NbHM Homo sapiens cDNA clone
            IMAGE:269265 5', mRNA sequence.
ACCESSION   N39807
VERSION     N39807.1  GI:1163352
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 356
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: T7
            High quality sequence stop: 356.
FEATURES             Location/Qualifiers
     source          1. .410
                     /organism="Homo sapiens"
                     /db_xref="GDB:3878907"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:269265"
                     /clone_lib="Soares melanocyte 2NbHM"
                     /sex="Male"
                     /tissue_type="melanocyte"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCAGTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library constructed by Bento Soares and
                     M.Fatima Bonaldo. RNA from normal foreskin melanocytes
                     (FS374) was kindly provided by Dr. Anthony P. Albino."
BASE COUNT      127 a     88 c     72 g    122 t      1 others
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x N39807   ..

 Align seg 1/1  to: N39807  from: 1  to: 410

      11 SerAlaLeuLeuThrAlaLeuLeu 18
         ||||||||||||||||||||||||
      91 TCAGCGCTTCTTACAGCTTTGCTC 114
```

```
seq_name: gb_est2:BJ072077

seq_documentation_block:
LOCUS       BJ072077                411 bp    mRNA    linear   EST 11-DEC-2001
DEFINITION  BJ072077 NIBB Mochii normalized Xenopus tailbud library Xenopus
            laevis cDNA clone XL096h02 5', mRNA sequence.
ACCESSION   BJ072077
VERSION     BJ072077.1  GI:17502266
KEYWORDS    EST.
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 411)
  AUTHORS   Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
            ,Y.
  TITLE     Expressed genes in X. laevis embryo
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .411
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /clone="XL096h02"
                     /clone_lib="NIBB Mochii normalized Xenopus tailbud
                     library"
                     /tissue_type="whole embryo"
                     /dev_stage="stage 25"
BASE COUNT      126 a     85 c    107 g     91 t      2 others
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BJ072077/rev   ..

 Align seg 1/1  to reverse of: BJ072077  from: 1  to: 411

      11 SerAlaLeuLeuThrAlaLeuLeu 18
         ||||||||||||||||||||||||
     260 TCTGCTCTTCTCACAGCACTGCTG 237

seq_name: gb_gss:AZ132522

seq_documentation_block:
LOCUS       AZ132522                427 bp    DNA     linear   GSS 02-JUN-2000
DEFINITION  OSJNBb0070H11f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone OSJNBb0070H11f, DNA sequence.
ACCESSION   AZ132522
VERSION     AZ132522.1  GI:8211144
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GTAAAACGACGGCCAGTG
            Class: BAC ends
            High quality sequence start: 23
            High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="OSJNBb0070H11f"
                     /clone_lib="CUGI Rice BAC Library (EcoRI)"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                     Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Monocotyledonous rice is a diploid plant (2n=24) with a
                     haploid genome equivalent of 431 Mbp (Arumuganathan and
                     Earle, 1991). The relatively small genome of rice, three
                     times larger than that of Arabidopsis, makes it suitable
                     for genomic studies. In order to facilitate positional
                     cloning, physical mapping and genome sequencing of rice,
                     we have constructed a BAC library from Oryza sativa,
                     Nipponbare variety using EcoRI as the cloning enzyme. The
                     library contains 55,296 clones with an average insert size
                     of 121 Kb providing approximatley 15 haploid genome
                     equivalents. The deep coverage allows the isolation a
                     particular sequence with a probability of 99.9 %. Three
                     high density filters, each containing 18,432 clones
                     (doubly spotted), represent the whole library for colony
                     screening and can be requested from the Clemson University
                     BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT      126 a    102 c     90 g    109 t
ORIGIN


alignment_scores:
            Quality:    8.00            Length:      8
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AZ132522   ..

 Align seg 1/1  to: AZ132522  from: 1  to: 427

      15 ThrAlaLeuLeuValThrGlyCys 22
         ||||||||||||||||||||||||
     334 ACTGCATTGCTGGTCACTGGCTGC 357

seq_name: gb_est2:BE685620

seq_documentation_block:
LOCUS       BE685620                430 bp    mRNA    linear   EST 11-SEP-2000
DEFINITION  uu56d10.x1 Soares_thymus_2NbMT Mus musculus cDNA clone
            IMAGE:3375955 3', mRNA sequence.
ACCESSION   BE685620
VERSION     BE685620.1  GI:10073296
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
```

```
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL    Unpublished (1997)
COMMENT      Other_ESTs: uu56d10.y1
             Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:1085559.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3375955"
                     /clone_lib="Soares_thymus_2NbMT"
                     /sex="male"
                     /tissue_type="Thymus"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGTTTTTTTTTTTTTTTTTTTTTTT
                     3']; double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. RNA
                     provided by Dr. Bertrand Jordan. Library went through two
                     rounds of normalization, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      153 a     90 c     77 g    110 t
ORIGIN


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x BE685620/rev   ..

 Align seg 1/1  to reverse of: BE685620  from: 1  to: 430

      22 CysValSerThrGlyAsnValAla 29
         ||||||||||||||||||||||||
     236 TGTGTCAGTACTGGAAATGTGGCC 213
```

```
OM of: US-10-048-197-2 to: N_Geneseq_032802:*   out_format : pfs

Date: Sep 18, 2002 10:24 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US10048197/runat_17092002_135957_23607/app_query.fasta_1.168
 -DB=N_Geneseq_032802 -QFMT=fastap -SUFFIX=oli.rng -GAPOP=4.500
 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
 -THR_MIN=1 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US10048197_@CGN1_1_523 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-10-048-197-2
Query length: 111
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 858457221
Search time (sec): 277.840000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000


score_list:
Sequence            Strd Orig    %Score   EScore Len  ! Documentation  ..
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF59778 +  111.00 1992.65  1.2e-102  3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF28551 +  111.00 1951.02  2.5e-100  9
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH65480 +    8.00  137.29  26.74     2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI89885 +    8.00  132.81  47.50     4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX13630 +    8.00  132.75  47.88     48
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABK16762 -    8.00  131.63  55.26     56
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH05525 +    8.00  129.41  73.44     7
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:AAQ62008 -    8.00  128.28  84.93     88
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK53351 +    8.00  128.07  87.28     9
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC39230 +    8.00  126.48  107.02    11
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS81888 +    8.00  125.18  126.38    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH15929 +    8.00  120.14  241.23    2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK52367 +    8.00  118.45  299.58    3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF24443 +    8.00  116.50  384.85    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA16230 +    8.00  116.11  404.42    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA16231 +    8.00  116.11  404.58    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA16232 +    8.00  116.07  406.33    4
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL05316 -    8.00  111.69  712.60    8
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA18955 +    8.00  110.29  852.67    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA19362 -    8.00  110.29  852.67    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAL36875 -    8.00  108.94  1.0e+03   1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA19852 +    8.00  103.35  2.1e+03   2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS36105 -    8.00  101.76  2.5e+03   3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS31532 -    8.00  101.76  2.5e+03   3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS31538 -    8.00  101.76  2.5e+03   3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL04952 +    8.00   98.10  4.1e+03   5
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF28546 +    8.00   96.74  4.9e+03   6
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH41226 -    8.00   84.15  2.4e+04   3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH68525 -    8.00   84.15  2.4e+04   3
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA72068 +    7.00  125.31  124.35    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK20495 +    7.00  125.31  124.35    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK46634 +    7.00  125.31  124.35    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI52470 +    7.00  125.31  124.35    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC98752 +    7.00  122.78  171.99    16
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS48479 +    7.00  122.04  189.02    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS48487 +    7.00  121.49  202.99    1
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH84463 -    7.00  121.22  209.95    2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAC82842 +    7.00  119.58  259.31    2
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA16416 -    7.00  119.04  277.93
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA16419 -    7.00  119.04  277.93
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAA69628 -    7.00  118.26  307.17
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAZ12563 +    7.00  118.23  308.15
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH53149 +    7.00  117.73  328.52
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS33704 +    7.00  117.71  329.49
/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH32183 +    7.00  117.46  340.13

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF59778

seq_documentation_block:
ID   AAF59778 standard; DNA; 336 BP.
XX
AC   AAF59778;
XX
DT   02-MAY-2001  (first entry)
XX
DE   Moraxella catarrhalis strain ATCC43617 BASB122 DNA.
XX
KW   BASB122 protein; strain ATCC43617; antigen; antibody; vaccine;
KW   genetic immunisation; infection; upper respiratory tract; otitis media;
KW   hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
KW   invasive disease; antibacterial; auditory; ds.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200109337-A2.
XX
PD   08-FEB-2001.
XX
PF   31-JUL-2000; 2000WO-EP07365.
XX
PR   30-JUL-1999;   99GB-0018034.
PR   30-JUL-1999;   99GB-0018036.
XX
PA   (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI   Thonnard J;
XX
DR   WPI; 2001-159874/16.
DR   P-PSDB; AAB60640.
XX
PT   New BASB122 and BASB124 polypeptides and polynucleotides from Moraxella
PT   catarrhalis strain ATCC 43617, useful as therapeutic agents or vaccines
PT   against bacterial infections, e.g. otitis media or pneumonia
XX
PS   Claim 13; Page 66; 75pp; English.
XX
CC   The invention relates to the Moraxella catarrhalis strain ATCC43617
CC   BASB122 and BASB124 proteins (AAB60640 and AAB60641, respectively)
CC   and to DNA encoding them (AAF59778 and AAF59779, respectively). The
CC   invention also relates to immunogenic fragments of the BASB122 and
CC   BASB124 proteins, expression vectors and host cells comprising BASB122
CC   or BASB124 nucleic acids, the recombinant production of BASB122 or
CC   BASB124, vaccine compositions comprising the BASB122 or BASB124 proteins
CC   or nucleic acids, an antibody against BASB122 or BASB124, therapeutic
CC   compositions comprising an anti-BASB122 or anti-BASB124 antibody, and a
CC   method of identifying a Moraxella  catarrhalis infection via the
CC   detection of BASB122 or BASB124 proteins or antibodies. The vaccine
CC   compositions of the invention are useful as prophylactic or therapeutic
CC   agents against Moraxella catarrhalis infections in mammals, particularly
CC   humans. Moraxella catarrhalis is a Gram negative bacterium frequently
CC   isolated from the human upper respiratory tract, which is responsible for
CC   several pathological conditions. It is responsible for about 15% of
CC   otitis media cases in children (which can lead to temporary or permanent
CC   hearing loss). It also causes pneumonia in elderly people, and sinusitis,
CC   nosocomial infections and, less frequently, invasive diseases. BASB122 or
CC   BASB124 proteins or nucleotides may additionally be used in screening for
CC   novel antibacterial compounds, and in the diagnosis and staging of
CC   infections. The present sequence represents DNA encoding the Moraxella
CC   catarrhalis strain ATCC43617 BASB122 protein.
XX
SQ   Sequence 336 BP; 106 A; 65 C; 61 G; 104 T; 0 other;
```

```
alignment_scores:
            Quality:  111.00              Length:    111
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAF59778   ..

 Align seg 1/1  to: AAF59778  from: 1  to: 336

       1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
       1 ATGTTAAATCATCATATTCGCCTGACTATTTCCGCTTTACTGACGGCACT 50

      17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 TTTGGTAACAGGTTGTGTTTCTACTGGTAATGTTGCAATGAAAGAGCAAA 100

      34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 ACCAACAAACCATTGAGCAGACCATCATTAAGGGCAAGACCAATAAGCAA 150

      51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVa 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 GAGATTTCTAGTAGATTTGGTTCTGCTGATAGCATCTCTTTTATGATAGT 200

      67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 GGTAATAAAATTTGGACATACCGCCATACTCGCTCCAAACCGATGGCAAG 250

      84 luIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 AAATTTTATCCCTTATAATTTCTTTTCTTTGGGTGAAACCATACAGACCA 300

     101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
         |||||||||||||||||||||||||||||||||
     301 AAGAACTTGTCATTTTATTTGACAGCAAAGGCA 333

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF28551

seq_documentation_block:
ID   AAF28551 standard; DNA; 94750 BP.
XX
AC   AAF28551;
XX
DT   04-APR-2001  (first entry)
XX
DE   Genomic fragment #38.
XX
KW   Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW   bronchopulmonary; endocarditis; meningitis; ss.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200078968-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US16649.
XX
PR   18-JUN-1999;   99US-0140121.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lagace RE,  Patterson C,  Berg KL;
XX
DR   WPI; 2001-041427/05.
XX
PT   Genomic library for identifying diagnostic and therapeutic
PT   compositions, and for identifying virulence factors, regulatory
PT   elements and drug targets, comprises Moraxella catarrhalis nucleic
PT   acids -
XX
PS   Claim 1; Page 415-436; 545pp; English.
XX
CC   The present invention relates to a Moraxella catarrhalis genomic library
CC   comprising of a combination of 41 nucleic acid molecules (see
CC   AAF28514-AAF28554). The library has a number of uses described in the
CC   specification e.g. is useful for aidentifying diagnostic and therapeutic
CC   compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC   aerobic, gram-negative diplococcus, normally found among the bacterial
CC   flora of human upper airways. M. catarrhalis is known to cause acute,
CC   localised infections such as otitis media, sinusitis and bronchopulmonary
CC   infection and life-threatening, systemic diseases including endocarditis
CC   and meningitis.
XX
SQ   Sequence 94750 BP; 27128 A; 18811 C; 21444 G; 27367 T; 0 other;


alignment_scores:
            Quality:  111.00              Length:    111
              Ratio:  1.000                 Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAF28551   ..

 Align seg 1/1  to: AAF28551  from: 1  to: 94750

       1 MetLeuAsnHisHisIleArgLeuThrIleSerAlaLeuLeuThrAlaLe 17
         |||||||||||||||||||||||||||||||||||||||||||||||||||
   75267 ATGTTAAATCATCATATTCGCCTGACTATTTCCGCTTTACTGACGGCACT 75316

      17 uLeuValThrGlyCysValSerThrGlyAsnValAlaMetLysGluGlnA 34
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75317 TTTGGTAACAGGTTGTGTTTCTACTGGTAATGTTGCAATGAAAGAGCAAA 75366

      34 snGlnGlnThrIleGluGlnThrIleIleLysGlyLysThrAsnLysGln 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75367 ACCAACAAACCATTGAGCAGACCATCATTAAGGGCAAGACCAATAAGCAA 75416

      51 GluIleSerSerArgPheGlySerAlaAspSerIleSerPheMetIleVa 67
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75417 GAGATTTCTAGTAGATTTGGTTCTGCTGATAGCATCTCTTTTATGATAGT 75466

      67 lValIleLysPheGlyHisThrAlaIleLeuAlaProAsnArgTrpGlnG 84
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75467 GGTAATAAAATTTGGACATACCGCCATACTCGCTCCAAACCGATGGCAAG 75516

      84 luIleLeuSerLeuIleIleSerPheLeuTrpValLysProTyrArgPro 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
   75517 AAATTTTATCCCTTATAATTTCTTTTCTTTGGGTGAAACCATACAGACCA 75566

     101 LysAsnLeuSerPheTyrLeuThrAlaLysAla 111
         |||||||||||||||||||||||||||||||||
   75567 AAGAACTTGTCATTTTATTTGACAGCAAAGGCA 75599

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH65480

seq_documentation_block:
ID   AAH65480 standard; DNA; 261 BP.
XX
AC   AAH65480;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 515.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
```

```
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
DR   P-PSDB; AAG90261.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT    -
XX
PS   Claim 8; SEQ ID NO: 515; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 261 BP; 52 A; 77 C; 54 G; 78 T; 0 other;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAH65480   ..

 Align seg 1/1  to: AAH65480  from: 1  to: 261

      11 SerAlaLeuLeuThrAlaLeuLeu 18
         ||||||||||||||||||||||||
      95 TCCGCTTTGCTCACAGCCCTTCTT 118

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI89885

seq_documentation_block:
ID   AAI89885 standard; cDNA; 479 BP.
XX
AC   AAI89885;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 9945.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO09954.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 9945; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 479 BP; 103 A; 97 C; 102 G; 176 T; 1 other;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAI89885   ..

 Align seg 1/1  to: AAI89885  from: 1  to: 479

      86 LeuSerLeuIleIleSerPheLeu 93
         ||||||||||||||||||||||||
     348 CTAAGCTTAATCATTTCTTTTCTT 371

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:AAX13630

seq_documentation_block:
ID   AAX13630 standard; DNA; 483 BP.
XX
AC   AAX13630;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:693.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
```

XX
PF 04-MAY-1998; 98WO-US08985.
XX
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Dillon PJ, Kunsch CA;
XX
DR WPI; 1999-045171/04.
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
PS Claim 1; Page 1931-1932; 2084pp; English.
XX
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
SQ Sequence 483 BP; 140 A; 84 C; 97 G; 161 T; 1 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x AAX13630 ..

Align seg 1/1 to: AAX13630 from: 1 to: 483

7 ArgLeuThrIleSerAlaLeuLeu 14
||||||||||||||||||||||||
65 CGTTTAACAATTTCAGCACTTTTA 88

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:ABK16762

seq_documentation_block:
ID ABK16762 standard; cDNA; 562 BP.
XX
AC ABK16762;
XX
DT 14-MAR-2002 (first entry)
XX
DE Human cDNA encoding G protein-coupled receptor nGPCR-Seq2664.
XX
KW Human; ss; G protein-coupled receptor; nGPCR-x; thyroid disorder;
KW thyreotoxicosis; renal failure; inflammatory condition; Crohn's disease;
KW rheumatoid arthritis; autoimmune disorder; movement disorders;
KW CNS disorder; migraine; neurological disorder; anxiety;
KW manic depression; schizophrenia; epilepsy; severe mental retardation;
KW dyskinesia; Huntington's disease; Tourette's syndrome;
KW degenerative disorder; Parkinson's disease; Alzheimer's disease;
KW viral infection; HIV infection; cardiovascular disorder;
KW human immunodeficiency virus; metabolic disorder; type II diabetes;
KW obesity; anorexia; hypotension; hypertension; myocardial infarction;
KW atherosclerosis; cancer; hyperproliferative disorder; psoriasis;
KW hormonal disorder; polycystic ovarian syndrome; sexual dysfunction;
KW respiratory ailment; asthma; traumatic brain injury; angina;
KW glucose uptake disorder; metabolic disorder; angiogenesis;
KW vasculogenesis; wound healing.
XX
OS Homo sapiens.
XX
PN WO200190149-A2.
XX
PD 29-NOV-2001.
XX
PF 22-MAY-2001; 2001WO-US16419.
XX
PR 22-MAY-2000; 2000US-206138P.
PR 22-MAY-2000; 2000US-206139P.
PR 02-JUN-2000; 2000US-208976P.
XX
PA (PHAA) PHARMACIA & UPJOHN CO.
XX
PI Vogeli G;
XX
DR WPI; 2002-106173/14.
DR P-PSDB; AAU80550.
XX
PT Novel G-protein-coupled receptor polypeptides and polynucleotides
PT useful for diagnosis, prevention and treatment of viral infections,
PT cancer, metabolic, cardiovascular diseases and neurological disorders -
XX
PS Claim 4; Page 71; 138pp; English.
XX
CC The invention relates to novel G-protein-coupled receptor
CC polynucleotides, termed nGPCR-x, and the nGPCR-x proteins. The nGPCR-x
CC proteins, polynucleotides, anti-nGPCR-x antibodies, and modulators of
CC nGPCR-x are useful for the diagnosis and treatment of diseases or
CC conditions such as thyroid disorders (thyreotoxicosis), renal failure,
CC inflammatory conditions (Crohn's disease), disease related to cell
CC differentiation and homeostasis, rheumatoid arthritis,
CC autoimmune disorders, movement disorders, CNS disorders (e.g. pain
CC including migraine, stroke, psychotic and neurological disorders
CC including anxiety, mental disorders, manic depression, generalised
CC anxiety disorders, post-traumatic-stress disorders, schizophrenia,
CC depression, dipolar disorder, delirium, epilepsy, dementia, severe mental
CC retardation, dyskinesias such as Huntington's disease or Tourette's
CC syndrome, attention disorders and degenerative disorders such as
CC Parkinson's, Alzheimer's disease, movement disorders including ataxias,
CC supranuclear palsy), infections such as viral infections caused by HIV-1
CC or HIV-2 (human immunodeficiency virus), metabolic and cardiovascular
CC disease and disorders (e.g. type II diabetes, impaired glucose tolerance,
CC dyslipidaemia, obesity, anorexia, hypotension, hypertension, thrombosis,
CC myocardial infarction, cardiomyopathies, atherosclerosis), proliferative
CC disease and cancers (e.g. breast, colon, lung cancer and
CC hyperproliferative disorders such as psoriasis, prostate hyperplasia),
CC hormonal disorders (male/female hormonal replacement, polycystic ovarian
CC syndrome, alopecia), sexual dysfunction, respiratory ailments such
CC as asthma, acute and/or traumatic brain injury and unstable angina due
CC to the role of T cells, inflammation, irregularities in glucose uptake
CC and metabolism, fatty acid uptake, abnormal angiogenesis, vasculogenesis,
CC wound healing. The present sequence is a cDNA encoding a nGPCR-x.
XX
SQ Sequence 562 BP; 176 A; 118 C; 140 G; 128 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x ABK16762/rev ..

Align seg 1/1 to reverse of: ABK16762 from: 1 to: 562

```
      10 IleSerAlaLeuLeuThrAlaLeu 17
         ||||||||||||||||||||||||
      66 ATTTCTGCGCTTCTCACTGCCTTG 43

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH05525

seq_documentation_block:
ID   AAH05525 standard; cDNA; 759 BP.
XX
AC   AAH05525;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA clone (5'-primer) SEQ ID NO:2360.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 1; SEQ ID 2360;.2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 759 BP; 184 A; 213 C; 210 G; 149 T; 3 other;


alignment_scores:
           Quality:    8.00              Length:      8
             Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAH05525   ..

 Align seg 1/1  to: AAH05525  from: 1  to: 759

      12 AlaLeuLeuThrAlaLeuLeuVal 19
         |||||||||||||||||||||||||
      63 GCCCTGCTAACAGCACTGCTGGTA 86

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:AAQ62008

seq_documentation_block:
ID   AAQ62008 standard; cDNA to mRNA; 885 BP.
XX
AC   AAQ62008;
XX
DT   26-SEP-1994  (first entry)
XX
DE   SSO2 gene.
XX
KW   Sec1 suppressor gene; homologue; secretion; biomass; yeast; ss.
XX
OS   Saccharomyces cerevisiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..885
FT                   /*tag=  a
XX
PN   WO9408024-A.
XX
PD   14-APR-1994.
XX
PF   06-OCT-1993;   93WO-FI00402.
XX
PR   06-OCT-1992;   92FI-0004494.
XX
PA   (VALW ) VALTION TEKNILLINEN TUTKIMUSKESKUS.
XX
PI   Aalto M,  Keraenen S,  Outola M,  Penttilae M,  Ronne H;
XX
DR   WPI; 1994-135588/16.
DR   P-PSDB; AAR51274.
XX
PT   New sec1 suppressor gene - used for increasing prodn of secreted
PT   foreign or endogenous proteins by eukaryotic cells
XX
PS   Claim 3; Page 32; 51pp; English.
XX
CC   The yeast sec1 suppressor gene SSO2 was isolated as a supreesor of
CC   the temperature sensitive defect of the sec 1-1 mutant.  When
CC   overexpressed in eukaryotic cells the gene renders the cells capable
CC   of producing increased amts. of secreted proteins, e.g. antibodies
CC   or hydrolytic enzymes.  The gene may be used for more efficient
CC   biomass prodn. from raw material or efficient hydrolysis of a raw
CC   material.
CC   See also AAQ62007.
XX
SQ   Sequence 885 BP; 320 A; 171 C; 196 G; 198 T; 0 other;


alignment_scores:
           Quality:    8.00              Length:      8
             Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAQ62008/rev   ..

 Align seg 1/1  to reverse of: AAQ62008  from: 1  to: 885
```

```
      12 AlaLeuLeuThrAlaLeuLeuVal 19
         |||||||||||||||||||||||
     779 GCACTCTTAACGGCCTTGTTGGTG 756

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK53351

seq_documentation_block:
ID   AAK53351 standard; cDNA; 911 BP.
XX
AC   AAK53351;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 2880.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM80218.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 1; Page 5035; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence 911 BP; 205 A; 273 C; 244 G; 189 T; 0 other;


alignment_scores:
           Quality:    8.00            Length:      8
             Ratio:  1.000              Gaps:      0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAK53351   ..

 Align seg 1/1  to: AAK53351  from: 1  to: 911

      12 AlaLeuLeuThrAlaLeuLeuVal 19
         |||||||||||||||||||||||
     178 GCCCTGCTAACAGCACTGCTGGTA 201

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC39230

seq_documentation_block:
ID   AAC39230 standard; DNA; 1130 BP.
XX
AC   AAC39230;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 23849.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
```

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment_block:
 US-10-048-197-2 x AAC39230   ..

 Align seg 1/1  to: AAC39230  from: 1  to: 1130

      13 LeuLeuThrAlaLeuLeuValThr 20
         ||||||||||||||||||||||||
     637 CTTTTAACCGCTTTGCTCGTGACA 660

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS81888

seq_documentation_block:
ID   AAS81888 standard; cDNA; 1347 BP.
XX
AC   AAS81888;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #17692.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG17701.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 17692; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1347 BP; 274 A; 331 C; 346 G; 396 T; 0 other;

alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAS81888   ..

 Align seg 1/1  to: AAS81888  from: 1  to: 1347

      13 LeuLeuThrAlaLeuLeuValThr 20
         ||||||||||||||||||||||||
    1100 TTGCTGACTGCATTGTTGGTTACC 1123

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH15929

seq_documentation_block:
ID   AAH15929 standard; cDNA; 2667 BP.
XX
AC   AAH15929;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:14503.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14503; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
```

```
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2667 BP; 571 A; 655 C; 760 G; 681 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAH15929   ..

 Align seg 1/1  to: AAH15929  from: 1  to: 2667

      12 AlaLeuLeuThrAlaLeuLeuVal 19
         ||||||||||||||||||||||||
      63 GCCCTGCTAACAGCACTGCTGGTA 86

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK52367

seq_documentation_block:
ID   AAK52367 standard; cDNA; 3353 BP.
XX
AC   AAK52367;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 912.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM79234.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 1; Page 3005-3007; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence 3353 BP; 731 A; 802 C; 935 G; 885 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAK52367   ..

 Align seg 1/1  to: AAK52367  from: 1  to: 3353

      12 AlaLeuLeuThrAlaLeuLeuVal 19
         ||||||||||||||||||||||||
     221 GCCCTGCTAACAGCACTGCTGGTA 244

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF24443

seq_documentation_block:
ID   AAF24443 standard; DNA; 4369 BP.
XX
AC   AAF24443;
XX
DT   08-MAY-2001  (first entry)
XX
DE   Corynebacterium thermoaminogenes plasmid pYM2.
XX
KW   Plasmid; pYM2; rep; coryneform bacterium; L-amino acid; thermostable;
KW   high temperature; ds.
XX
OS   Corynebacterium thermoaminogenes.
XX
PN   EP1076094-A2.
XX
PD   14-FEB-2001.
XX
PF   11-AUG-2000; 2000EP-0117225.
XX
PR   12-AUG-1999;   99JP-0228391.
XX
PA   (AJIN ) AJINOMOTO CO INC.
XX
PI   Matsuzaki Y,  Kimura E,  Nakamatsu T,  Kurahashi O,  Kawahara Y;
PI   Sugimoto S;
XX
DR   WPI; 2001-193202/20.
DR   P-PSDB; AAB24440.
XX
PT   New plasmids derived from Corynebacterium thermoaminogenes, useful for
PT   improving coryneform bacteria, which can grow at elevated temperatures,
PT   and for producing useful substances (e.g. L-amino acids)  -
XX
PS   Example 2; Page 21-24; 32pp; English.
XX
CC   The present invention provides plasmids isolated from Corynebacterium
CC   thermoaminogenes encoding a rep protein. These are useful for improving
CC   Coryneform bacteria, which are capable of growing at elevated
CC   temperatures and of producing substances such as L-amino acids. The
CC   present sequence is the coding sequence of the Coryneform bacterial
CC   plasmid pYM2.
XX
```

```
SQ   Sequence 4369 BP; 965 A; 1252 C; 1293 G; 859 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAF24443    ..

 Align seg 1/1  to: AAF24443  from: 1  to: 4369

      51 GluIleSerSerArgPheGlySer 58
         ||||||||||||||||||||||||
    2010 GAAATTTCCTCTCGCTTCGGTAGT 2033

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA16230

seq_documentation_block:
ID   ABA16230 standard; DNA; 4604 BP.
XX
AC   ABA16230;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 8561.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   20-OCT-2000; 2000US-0242221.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
```

```
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 8561; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4604 BP; 1193 A; 1220 C; 1268 G; 923 T; 0 other;


alignment_scores:
           Quality:    8.00            Length:      8
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000


alignment_block:
 US-10-048-197-2 x ABA16230   ..

 Align seg 1/1  to: ABA16230  from: 1  to: 4604

      17 LeuLeuValThrGlyCysValSer 24
         ||||||||||||||||||||||||
     592 CTCCTGGTCACAGGCTGCGTTTCT 615

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA16231

seq_documentation_block:
ID   ABA16231 standard; DNA; 4606 BP.
XX
AC   ABA16231;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 8562.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
```

PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 8562; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4606 BP; 1195 A; 1223 C; 1267 G; 921 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x ABA16231 ..

Align seg 1/1 to: ABA16231 from: 1 to: 4606

17 LeuLeuValThrGlyCysValSer 24

```
          ||||||||||||||||||||||||
     592 CTCCTGGTCACAGGCTGCGTTTCT 615

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA16232

seq_documentation_block:
ID   ABA16232 standard; DNA; 4627 BP.
XX
AC   ABA16232;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 8563.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   20-OCT-2000; 2000US-0242221.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
```

```
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 8563; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 4627 BP; 1191 A; 1234 C; 1282 G; 920 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x ABA16232    ..

 Align seg 1/1  to: ABA16232  from: 1  to: 4627

      17 LeuLeuValThrGlyCysValSer 24
         ||||||||||||||||||||||||
     588 CTCCTGGTCACAGGCTGCGTTTCT 611

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL05316

seq_documentation_block:
ID   ABL05316 standard; cDNA; 8377 BP.
XX
AC   ABL05316;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 10430.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB61213.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 10430; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 8377 BP; 2677 A; 1829 C; 1525 G; 2346 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x ABL05316/rev    ..

 Align seg 1/1  to reverse of: ABL05316  from: 1  to: 8377

      34 AsnGlnGlnThrIleGluGlnThr 41
         |||||||||||||||||||||||
    2698 AATCAACAAACAATCGAACAAACA 2675

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA18955

seq_documentation_block:
ID   ABA18955 standard; DNA; 10126 BP.
XX
AC   ABA18955;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 11286.
```

XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.

```
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 11286; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 10126 BP; 2333 A; 2248 C; 2796 G; 2749 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x ABA18955   ..

 Align seg 1/1  to: ABA18955  from: 1  to: 10126

      49 LysGlnGluIleSerSerArgPhe 56
         ||||||||||||||||||||||||
    8429 AAGCAAGAAATATCAAGTAGGTTT 8452

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA19362

seq_documentation_block:
ID   ABA19362 standard; DNA; 10126 BP.
XX
AC   ABA19362;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 11693.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
```

```
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   20-OCT-2000; 2000US-0242221.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250391.
PR   01-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 11693; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 10126 BP; 2749 A; 2796 C; 2248 G; 2333 T; 0 other;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x ABA19362/rev   ..

 Align seg 1/1  to reverse of: ABA19362  from: 1  to: 10126

      49 LysGlnGluIleSerSerArgPhe 56
         |||||||||||||||||||||||
    1698 AAGCAAGAAATATCAAGTAGGTTT 1675

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAL36875

seq_documentation_block:
ID   AAL36875 standard; DNA; 12167 BP.
XX
AC   AAL36875;
XX
DT   08-JAN-2002  (first entry)
XX
DE   Human musculoskeletal system related polynucleotide SEQ ID NO 3240.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein;
KW   musculoskeletal system; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155367-A1.
XX
PD   02-AUG-2001.
XX
```

PF 17-JAN-2001; 2001WO-US01338.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451937/48.
XX
PT Isolated polypeptide for treating, preventing and/ or prognosing

```
PT   disorders related to the musculoskeletal system including
PT   musculoskeletal cancers and also for testing and detection e.g.
PT   diagnosis -
XX
PS   Example 2; SEQ ID NO 3240; 781pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC   (ABB03087-ABB04109) associated with the musculoskeletal system useful
CC   for preventing, treating or ameliorating medical conditions e.g. by
CC   protein or gene therapy. The genes are isolated from a range of human
CC   tissues disclosed in the specification. The nucleic acids, proteins,
CC   antibodies and (ant)agonists are useful in the diagnosis, treatment
CC   and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC   other cancers of the adrenal gland, bone, bone marrow, breast,
CC   gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC   (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC   healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC   and (f) infectious diseases such as viral, bacterial, fungal and
CC   parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 12167 BP; 3252 A; 2986 C; 2599 G; 3330 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAL36875/rev   ..

 Align seg 1/1  to reverse of: AAL36875  from: 1  to: 12167

      88 LeuIleIleSerPheLeuTrpVal 95
         ||||||||||||||||||||||||
   10917 CTGATAATTTCTTTCCTTTGGGTA 10894

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA19852

seq_documentation_block:
ID   ABA19852 standard; DNA; 25955 BP.
XX
AC   ABA19852;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 12183.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
```

PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 12183; 1701pp + Sequence Listing; English.
XX

CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 25955 BP; 8559 A; 4165 C; 4808 G; 8423 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x ABA19852 ..

Align seg 1/1 to: ABA19852 from: 1 to: 25955

12 AlaLeuLeuThrAlaLeuLeuVal 19
||||||||||||||||||||||||
16733 GCATTGCTCACTGCTTTGTTGGTA 16756

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS36105

seq_documentation_block:
ID AAS36105 standard; DNA; 32195 BP.
XX
AC AAS36105;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human cardiovascular system antigen genomic DNA SEQ ID No 1605.
XX
KW Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; tissue regeneration;
KW anti-infertility.
XX
OS Homo sapiens.
XX
PN WO200155321-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01340.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451930/48.
XX
PT New cardiovascular system related polynucleotides and polypeptides,
PT useful for diagnosing, treating and/or preventing disorders of the
PT cardiovascular system -
XX
PS Claim 1; SEQ ID No 1605; 674pp; English.
XX
CC Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC the cardiovascular system antigen polypeptides of the invention.
CC Cardiovascular system antigens and their associated polynucleotides are

```
CC   useful in the diagnosis, treatment and prevention of various types of
CC   disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC   chickens or sheep. A pathological condition can be determined by
CC   detecting the presence or absence of a mutation in a cardiovascular
CC   system antigen polynucleotide. The treatable disorders include autoimmune
CC   diseases such as rheumatoid arthritis, hyperproliferative disorders such
CC   as neoplasms of the breast or liver, cardiovascular disorders such as
CC   cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
CC   nervous system disorders such as Alzheimer's disease, infections caused
CC   by bacteria, viruses and fungi, ocular disorders such as corneal
CC   infection, endocrine disorders such as premature labour and infertility,
CC   gastrointestinal disorders such as Crohn's disease, renal disorders such
CC   as glomerulonephritis and respiratory disorders such as asthma and
CC   pleurisy. The polypeptides can also be used to aid wound healing, to
CC   prevent skin aging due to sunburn, to maintain organs before
CC   transplantation, to regenerate tissues and in chemotaxis.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX

alignment_scores:
           Quality:    8.00              Length:      8
             Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAS36105/rev   ..

 Align seg 1/1  to reverse of: AAS36105  from: 1  to: 32195

      85 IleLeuSerLeuIleIleSerPhe 92
         ||||||||||||||||||||||||
   25174 ATACTATCTTTAATAATATCATTT 25151

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS31532

seq_documentation_block:
ID   AAS31532 standard; DNA; 32195 BP.
XX
AC   AAS31532;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Human DNA for a novel extracellular matrix protein, Seq ID No 611.
XX
KW   Human; secreted extracellular matrix protein; ds; immunomodulatory;
KW   Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiant; vascular;
KW   cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;
KW   antialzheimers; immune/autoimmune disease; HIV infection; anaemia;
KW   human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW   cancers; hyperproliferative disorder; breast neoplasm; melanoma;
KW   Sezary syndrome; Gaucher's disease; neurological diseases;
KW   Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW   cardiac arrest; tachycardia; angina; infection; corneal infections;
KW   wound healing; immunogen; gene therapy; antisense; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200155368-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01348.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
```

```
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465572/50.
XX
PT   Nucleic acid molecules encoding human secreted extracellular matrix
PT   proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT   Alzheimer's and Parkinson's diseases and cancers -
XX
PS   Claim 1; SEQ ID No 611; 577pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules encoding
CC   novel human secreted extracellular matrix proteins (SPs). The
CC   polynucleotides and proteins are used to prevent, treat a medical
CC   condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC   chickens or sheep. For example, disorders associated with decreased
CC   expression of SPs. The SP polynucleotide or a vector expressing them may
CC   be administered to treat diseases by gene therapy. Antisense molecules
CC   may be administered to down regulate expression of SPs by binding with
CC   the cells own genes and preventing their expression. The polynucleotides
CC   may also be used as DNA probes in diagnostic assays. The SPs may also be
CC   used as antigens to produce antibodies and to identify modulators
CC   (agonists and antagonists) of the SPs. The anti-(SP) antibodies and
CC   antagonists may also be used to down regulate expression and activity of
CC   SP and as diagnostic agents for detecting the presence of SPs in samples.
CC   The disorders include for example: immune/autoimmune diseases (e.g. HIV
CC   (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC   and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC   melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC   Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC   Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac
CC   arrest, tachycardia and angina), infections caused by bacteria, viruses
CC   and fungi and ocular disorders (e.g. corneal infections). Other uses
CC   include wound healing, maintenance of organs before transplantation,
CC   support of cell culture of primary tissues, modulation of for example

alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAS31532/rev   ..

 Align seg 1/1  to reverse of: AAS31532  from: 1  to: 32195

      85 IleLeuSerLeuIleIleSerPhe 92
         |||||||||||||||||||||||
   25174 ATACTATCTTTAATAATATCATTT 25151

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAS31538

seq_documentation_block:
ID   AAS31538 standard; DNA; 32195 BP.
XX
AC   AAS31538;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Human DNA for a novel extracellular matrix protein, Seq ID No 617.
XX
KW   Human; secreted extracellular matrix protein; ds; immunomodulatory;
KW   Anti-HIV; antianemic; antirheumatic; antisclerotic; cardiant; vascular;
KW   cerebroprotective; thrombolytic; antimicrobial; ophthalmic; cytostatic;
KW   antialzheimers; immune/autoimmune disease; HIV infection; anaemia;
KW   human immunodeficiency virus; rheumatoid arthritis; multiple sclerosis;
KW   cancers; hyperproliferative disorder; breast neoplasm; melanoma;
KW   Sezary syndrome; Gaucher's disease; neurological diseases;
KW   Alzheimer's disease; Parkinson's disease; cardiovascular disorder;
KW   cardiac arrest; tachycardia; angina; infection; corneal infections;
KW   wound healing; immunogen; gene therapy; antisense; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200155368-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01348.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
```

PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465572/50.
XX
PT Nucleic acid molecules encoding human secreted extracellular matrix
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers -
XX
PS Claim 1; SEQ ID No 617; 577pp; English.
XX
CC The invention relates to isolated nucleic acid molecules encoding
CC novel human secreted extracellular matrix proteins (SPs). The
CC polynucleotides and proteins are used to prevent, treat a medical
CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. For example, disorders associated with decreased
CC expression of SPs. The SP polynucleotide or a vector expressing them may

```
CC   be administered to treat diseases by gene therapy. Antisense molecules
CC   may be administered to down regulate expression of SPs by binding with
CC   the cells own genes and preventing their expression. The polynucleotides
CC   may also be used as DNA probes in diagnostic assays. The SPs may also be
CC   used as antigens to produce antibodies and to identify modulators
CC   (agonists and antagonists) of the SPs. The anti-(SP) antibodies and
CC   antagonists may also be used to down regulate expression and activity of
CC   SP and as diagnostic agents for detecting the presence of SPs in samples.
CC   The disorders include for example: immune/autoimmune diseases (e.g. HIV
CC   (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis
CC   and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC   melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC   Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC   Parkinson's disease) cardio-/cerebrovascular disorders (e.g. cardiac
CC   arrest, tachycardia and angina), infections caused by bacteria, viruses
CC   and fungi and ocular disorders (e.g. corneal infections). Other uses
CC   include wound healing, maintenance of organs before transplantation,
CC   support of cell culture of primary tissues, modulation of for example

alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAS31538/rev   ..

 Align seg 1/1  to reverse of: AAS31538  from: 1  to: 32195

      85 IleLeuSerLeuIleIleSerPhe 92
         |||||||||||||||||||||||||
   25174 ATACTATCTTTAATAATATCATTT 25151

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:ABL04952

seq_documentation_block:
ID   ABL04952 standard; cDNA; 52872 BP.
XX
AC   ABL04952;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 9338.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB60849.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 9338; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 52872 BP; 15532 A; 11487 C; 10803 G; 15050 T; 0 other;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x ABL04952   ..

 Align seg 1/1  to: ABL04952  from: 1  to: 52872

      12 AlaLeuLeuThrAlaLeuLeuVal 19
         |||||||||||||||||||||||||
    9772 GCATTATTAACTGCTTTGCTTGTT 9795

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAF28546

seq_documentation_block:
ID   AAF28546 standard; DNA; 63563 BP.
XX
AC   AAF28546;
XX
DT   04-APR-2001  (first entry)
XX
DE   Genomic fragment #33.
XX
KW   Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW   bronchopulmonary; endocarditis; meningitis; ss.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200078968-A2.
XX
PD   28-DEC-2000.
XX
PF   16-JUN-2000; 2000WO-US16649.
XX
PR   18-JUN-1999;   99US-0140121.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lagace RE,  Patterson C,  Berg KL;
XX
DR   WPI; 2001-041427/05.
XX
PT   Genomic library for identifying diagnostic and therapeutic
PT   compositions, and for identifying virulence factors, regulatory
PT   elements and drug targets, comprises Moraxella catarrhalis nucleic
PT   acids -
XX
PS   Claim 1; Page 309-324; 545pp; English.
XX
CC   The present invention relates to a Moraxella catarrhalis genomic library
CC   comprising of a combination of 41 nucleic acid molecules (see
CC   AAF28514-AAF28554). The library has a number of uses described in the
CC   specification e.g. is useful for aidentifying diagnostic and therapeutic
CC   compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC   aerobic, gram-negative diplococcus, normally found among the bacterial
CC   flora of human upper airways. M. catarrhalis is known to cause acute,
CC   localised infections such as otitis media, sinusitis and bronchopulmonary
CC   infection and life-threatening, systemic diseases including endocarditis
```

```
CC   and meningitis.
XX
SQ   Sequence 63563 BP; 18440 A; 12630 C; 11985 G; 20508 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAF28546    ..

 Align seg 1/1  to: AAF28546  from: 1  to: 63563

      84 GluIleLeuSerLeuIleIleSer 91
         ||||||||||||||||||||||||
   45213 GAAATACTATCTCTTATTATCTCA 45236

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH41226

seq_documentation_block:
ID   AAH41226 standard; DNA; 349980 BP.
XX
AC   AAH41226;
XX
DT   29-OCT-2001  (first entry)
XX
DE   Pyrococcus abyssi genomic fragment #5.
XX
KW   Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX
OS   Pyrococcus abyssi.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1..49980
FT                   /*tag= a
FT                   /note= "This sequence overlaps with the 3' end of
FT                   AAH41225"
FT   misc_feature    300001..349980
FT                   /*tag= b
FT                   /note= "This sequence overlaps with the 5' end of
FT                   AAH41227"
XX
PN   FR2792651-A1.
XX
PD   27-OCT-2000.
XX
PF   21-APR-1999;   99FR-0005034.
XX
PR   21-APR-1999;   99FR-0005034.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
PA   (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI   Forterre P,  Thierry JC,  Prieur D,  Dietrich J,  Lecompte O;
PI   Querellou J,  Weissenbach J,  Saurin W,  Heilig R;
XX
DR   WPI; 2001-126236/14.
XX
PT   New nucleotide sequences isolated from Pyrococcus abyssi encode
PT   proteins useful in industry -
XX
PS   Claim 1; Page 511-606; 1657pp; French.
XX
CC   The present invention relates to the genomic sequence of Pyrococcus
CC   abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
CC   hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
CC   vents. The present sequence is a fragment of the genomic sequence of P.
CC   abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41225
CC   and the 3' end of this sequence overlaps with the 5' end of AAH41227. The
CC   proteins of the present invention have various potential industrial uses,
CC   since the proteins are stable at very high temperatures, some up to 110
CC   degrees centigrade.
CC   Note: This patent is in the same patent family as WO200065062, which
CC   contains additional sequences as shown in AAB99132-AAB99143,
CC   AAH75903-AAH75920 and AAG66436.
XX
SQ   Sequence 349980 BP; 98084 A; 80447 C; 77665 G; 93784 T; 0 other;


alignment_scores:
            Quality:    8.00              Length:      8
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAH41226/rev    ..

 Align seg 1/1  to reverse of: AAH41226  from: 1  to: 349980

       9 ThrIleSerAlaLeuLeuThrAla 16
         ||||||||||||||||||||||||
  104650 ACAATATCAGCCCTTTTAACGGCC 104627

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH68525

seq_documentation_block:
ID   AAH68525 standard; DNA; 349980 BP.
XX
AC   AAH68525;
XX
DT   26-SEP-2001  (first entry)
XX
DE   C glutamicum coding sequence fragment SEQ ID NO: 7060.
XX
KW   Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW   organic acid synthesis; ds.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999;   99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S,  Mizoguchi H,  Ando S,  Hayashi M,  Ochiai K,  Yokoi H;
PI   Tateishi N,  Senoh A,  Ikeda M,  Ozaki A;
XX
DR   WPI; 2001-376931/40.
XX
PT   Novel polynucleotides derived from Coryneform bacteria, for identifying
PT   mutation point of a gene, measuring expression of a gene, analysing
PT   expression profile or pattern of a gene and identifying homologous gene
PT   -
XX
PS   Disclosure; SEQ ID NO: 7060; 246pp + Sequence Listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or expression pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a nucleic acid described
CC   in the exemplification of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
```

```
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence 349980 BP; 80289 A; 91081 C; 97378 G; 81232 T; 0 other;


alignment_scores:
            Quality:    8.00             Length:      8
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAH68525/rev    ..

 Align seg 1/1  to reverse of: AAH68525  from: 1  to: 349980

      11 SerAlaLeuLeuThrAlaLeuLeu 18
         ||||||||||||||||||||||||
  181297 TCCGCTTTGCTCACAGCCCTTCTT 181274

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA72068

seq_documentation_block:
ID   ABA72068 standard; DNA; 115 BP.
XX
AC   ABA72068;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #20373.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00669.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human fetal liver -
XX
PS   Claim 4; SEQ ID NO 20373; 639pp + sequence listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for
CC   measuring human gene expression in a sample derived from human foetal
CC   liver. The single exon nucleic acid probes may be used for predicting,
CC   measuring and displaying gene expression in samples derived from human
CC   fetal liver. The present sequence is a single exon nucleic acid
CC   probe of the invention.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 115 BP; 26 A; 35 C; 14 G; 40 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x ABA72068    ..

 Align seg 1/1  to: ABA72068  from: 1  to: 115

      87 SerLeuIleIleSerPheLeu 93
         |||||||||||||||||||||
      38 TCTTTGATTATTTCCTTTCTC 58

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK20495

seq_documentation_block:
ID   AAK20495 standard; DNA; 115 BP.
XX
AC   AAK20495;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 20486.
XX
KW   Human; brain expressed exon; gene expression analysis; probe;
KW   microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW   epilepsy; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00667.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains -
XX
PS   Example 4; SEQ ID NO: 20486; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is one of the probes of the
CC   invention.
XX
SQ   Sequence 115 BP; 26 A; 35 C; 14 G; 40 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
```

```
US-10-048-197-2 x AAK20495   ..

Align seg 1/1  to: AAK20495  from: 1  to: 115

     87 SerLeuIleIleSerPheLeu 93
        |||||||||||||||||||||
     38 TCTTTGATTATTTCCTTTCTC 58

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAK46634

seq_documentation_block:
ID   AAK46634 standard; DNA; 115 BP.
XX
AC   AAK46634;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 21191.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 21191; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention.
XX
SQ   Sequence 115 BP; 26 A; 35 C; 14 G; 40 T; 0 other;


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAK46634   ..

 Align seg 1/1  to: AAK46634  from: 1  to: 115

      87 SerLeuIleIleSerPheLeu 93
         |||||||||||||||||||||
      38 TCTTTGATTATTTCCTTTCTC 58

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAI52470

seq_documentation_block:
ID   AAI52470 standard; DNA; 115 BP.
XX
AC   AAI52470;
XX
DT   17-OCT-2001  (first entry)
XX
DE   Probe #21156 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
XX
PS   Claim 25; SEQ ID No 21156; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
SQ   Sequence 115 BP; 26 A; 35 C; 14 G; 40 T; 0 other;


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAI52470   ..

 Align seg 1/1  to: AAI52470  from: 1  to: 115

      87 SerLeuIleIleSerPheLeu 93
         |||||||||||||||||||||
      38 TCTTTGATTATTTCCTTTCTC 58

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAC98752

seq_documentation_block:
ID   AAC98752 standard; cDNA; 162 BP.
XX
AC   AAC98752;
XX
DT   09-MAR-2001  (first entry)
XX
```

```
DE   Human colon cancer antigen nucleotide sequence SEQ ID NO:762.
XX
KW   Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW   identification; cytostatic; cardioactive; neuroprotective; vulnerary;
KW   immunomodulatory; muscular; gynaecological; gastrointestinal;
KW   nephrotropic; antiinfective; antibacterial; gene therapy; wound;
KW   neural disorder; immune system disorder; muscular disorder;
KW   reproductive disorder; gastrointestinal disorder; renal disorder;
KW   infectious disease; cardiovascular disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055351-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05883.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587534/55.
DR   P-PSDB; AAB53995.
XX
PT   Colon cancer associated gene sequences, referred to as colon cancer
PT   antigens, useful for the treatment, prevention, and diagnosis of colon
PT   disorders such as colon cancer -
XX
PS   Claim 1; Page 1323; 2104pp; English.
XX
CC   AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC   called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC   human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC   neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC   vulnerary, nephrotropic, antiinfective and antibacterial activities, and
CC   can be used in gene therapy. The colon cancer antigen polynucleotides,
CC   proteins and antibodies to the proteins are useful for the prevention,
CC   treatment and diagnosis of colon disorders, such as colon cancer. The
CC   polynucleotides may be used in diagnostics and research, such as for
CC   chromosome identification, and as hybridisation probes. The proteins
CC   may also be used to prevent diseases such as neural disorders, immune
CC   system disorders, muscular disorders, reproductive disorders,
CC   gastrointestinal disorders, wounds, renal disorders, infectious
CC   diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC   AAB54007 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 162 BP; 34 A; 50 C; 33 G; 37 T; 8 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAC98752   ..

 Align seg 1/1  to: AAC98752  from: 1  to: 162

      13 LeuLeuThrAlaLeuLeuVal 19
         |||||||||||||||||||
      56 CTGCTNACAGCCTTACTTGTA 76

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS48479

seq_documentation_block:
ID   AAS48479 standard; DNA; 179 BP.
XX
AC   AAS48479;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Klebsiella pneumoniae cellular proliferation inhibitory sequence #3.
XX
KW   Antisense; ss; prokaryotic cellular proliferation;
KW   antibiotic; antibacterial; drug design.
XX
OS   Klebsiella pneumoniae.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 1; Seq ID No 1056; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence is an antisense
CC   oligonucleotide of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 179 BP; 68 A; 38 C; 36 G; 37 T; 0 other;


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAS48479   ..

 Align seg 1/1  to: AAS48479  from: 1  to: 179

       6 IleArgLeuThrIleSerAla 12
         ||||||||||||||||||||
       2 ATCAGGCTGACAATCAGCGCC 22
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:AAS48487

seq_documentation_block:
ID   AAS48487 standard; DNA; 193 BP.
XX
AC   AAS48487;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Klebsiella pneumoniae cellular proliferation inhibitory sequence #11.
XX
KW   Antisense; ss; prokaryotic cellular proliferation;
KW   antibiotic; antibacterial; drug design.
XX
OS   Klebsiella pneumoniae.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 1; Seq ID No 1064; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence is an antisense
CC   oligonucleotide of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 193 BP; 61 A; 47 C; 42 G; 43 T; 0 other;


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAS48487    ..

 Align seg 1/1  to: AAS48487  from: 1  to: 193

       6 IleArgLeuThrIleSerAla 12
         ||||||||||||||||||||
      55 ATCAGGCTGACAATCAGCGCC 75

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH84463

seq_documentation_block:
ID   AAH84463 standard; DNA; 200 BP.
XX
AC   AAH84463;
XX
DT   26-SEP-2001  (first entry)
XX
DE   E. coli growth and proliferation related DNA sequence SEQ ID NO:91.
XX
KW   Escherichia coli; growth; proliferation; microbial; antimicrobial;
KW   bacterial infection; microorganism; ds.
XX
OS   Escherichia coli.
XX
PN   WO200134810-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30950.
XX
PR   09-NOV-1999;   99US-0164415.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Forsyth RA,  Ohlsen K,  Zyskind J;
XX
DR   WPI; 2001-335933/35.
XX
PT   Novel nucleic acids that inhibit Escherichia coli proliferation, useful
PT   for screening for homologous genes and for designing expression vectors
PT   -
XX
PS   Claim 1; Page 123; 522pp; English.
XX
CC   AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
CC   related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
CC   growth and proliferation related proteins given in AAG99078 and AAG98830
CC   to AAG98999. (I) can be used as potential targets for the generation of
CC   new antimicrobial agents, and for identification of compounds which
CC   interact with the gene products of (I). In addition the expression of
CC   (I) and the purification of the proteins, the purified proteins can be
CC   used to generate reagents and screen small molecule libraries or other
CC   candidate compound libraries for compounds that can be further developed
CC   to yield novel antimicrobial compounds. In addition, nucleic acid probes
CC   complementary to (I) that are specific for particular species of
CC   microorganisms can be used to identify particular microorganism species
CC   in clinical specimens, therefore, providing a rapid and dependable
CC   method by which to identify the causative agents of a bacterial
CC   infection. Also, antibodies generated against proteins translated from
CC   mRNA transcribed from proliferation-required sequences can also be used
CC   to screen for specific microorganisms that produce such proteins in a
CC   species-specific manner. AAH84371 and AAH84670 represent sequencing
CC   primers used in the isolation of E. coli growth and proliferation
CC   related sequence, which are used in an example from the present
CC   invention.
XX
SQ   Sequence 200 BP; 52 A; 50 C; 49 G; 49 T; 0 other;


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
```

```
 US-10-048-197-2 x AAH84463/rev   ..

 Align seg 1/1  to reverse of: AAH84463  from: 1  to: 200

      11 SerAlaLeuLeuThrAlaLeu 17
         ||||||||||||||||||||
      56 TCCGCTCTGCTAACTGCCTTA 36

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAC82842

seq_documentation_block:
ID   AAC82842 standard; cDNA; 250 BP.
XX
AC   AAC82842;
XX
DT   20-MAR-2001  (first entry)
XX
DE   Murine LMTF clone 701252210H1 cDNA.
XX
KW   LMTF; lipid metabolism transcription factor; cytostatic; hepatotropic;
KW   antiarteriosclerotic; nephrotropic; cell proliferation disorder; cancer;
KW   atherosclerosis; cirrhosis; hepatitis; lipid disorder; murine; ss.
XX
OS   Mus musculus.
XX
PN   WO200073443-A1.
XX
PD   07-DEC-2000.
XX
PF   15-MAY-2000; 2000WO-US13393.
XX
PR   26-MAY-1999;   99US-0318978.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Yue H,  Kaser MR,  Baughn MR;
XX
DR   WPI; 2001-061530/07.
XX
PT   New polynucleotide encoding lipid metabolism transcription factor is
PT   useful as a probe for detecting mammalian nucleic acids in a sample
XX
PS   Claim 3; Page 46; 52pp; English.
XX
CC   This invention describes a novel purified mammalian nucleotide sequence
CC   (I) encoding a lipid metabolism transcription factor (LMTF) which has
CC   cytostatic, hepatotropic, antiarteriosclerotic and nephrotropic activity.
CC   The probe hybridizing (I) is useful for detecting a mammalian nucleic
CC   acid sequence in a sample. (I) is useful to screen a library of molecules
CC   to identify at least a molecule specific to (I), comprising combining (I)
CC   with a library of molecules under conditions of binding and detecting the
CC   specific binding. LMTF is useful for treating or preventing a condition
CC   associated with altered expression or activity of the protein, which
CC   includes cell proliferation disorders such as atherosclerosis, cirrhosis,
CC   hepatitis, cancer, and various lipid disorders.
XX
SQ   Sequence 250 BP; 56 A; 67 C; 48 G; 78 T; 1 other;


alignment_scores:
        Quality:    7.00            Length:      7
          Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x AAC82842   ..

 Align seg 1/1  to: AAC82842  from: 1  to: 250

      15 ThrAlaLeuLeuValThrGly 21
         ||||||||||||||||||||
      85 ACAGCACTGTTGGTAACCGGT 105

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA16416

seq_documentation_block:
ID   ABA16416 standard; DNA; 269 BP.
XX
AC   ABA16416;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 8747.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
```

PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 8747; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 269 BP; 22 A; 104 C; 23 G; 120 T; 0 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x ABA16416/rev ..

Align seg 1/1 to reverse of: ABA16416 from: 1 to: 269

87 SerLeuIleIleSerPheLeu 93
|||||||||||||||||||||||
31 TCCCTCATCATCAGTTTCCTC 11

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA16419

seq_documentation_block:
ID ABA16419 standard; DNA; 269 BP.
XX
AC ABA16419;
XX

DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 8750.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 8750; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 269 BP; 22 A; 104 C; 23 G; 120 T; 0 other;

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x ABA16419/rev ..

Align seg 1/1 to reverse of: ABA16419 from: 1 to: 269

87 SerLeuIleIleSerPheLeu 93
|||||||||||||||||||||
31 TCCCTCATCATCAGTTTCCTC 11

OM of: US-10-048-197-2 to: Issued_Patents_NA:* out_format : pfs

Date: Sep 18, 2002 10:19 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10048197/runat_17092002_135957_23562/app_query.fasta_1.168
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=oli.rni -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10048197_@CGN1_1_74 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-10-048-197-2
Query length: 111
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 77.900000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-411-706-3	-	8.00	126.20	24.29	885	S
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-318-978-15	+	7.00	117.67	72.50	250	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-747-221B-1	-	7.00	114.28	112.01	401	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-747-221B-3	+	7.00	114.28	112.01	401	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-005-051-1	-	7.00	114.28	112.01	401	S
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-005-051-3	+	7.00	114.28	112.01	401	S
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-367-198A-4	-	7.00	112.02	149.58	549	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-932-978-1	-	7.00	107.95	252.39	969	S
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-367-198A-1	-	7.00	107.94	252.63	970	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-961-083-7	-	7.00	107.83	256.23	985	S
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-684-862-9	-	7.00	105.66	338.54	1333	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-13	-	7.00	103.91	423.27	1699	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-21	-	7.00	103.69	435.64	1753	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-31	-	7.00	103.68	436.33	1756	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-23	-	7.00	103.60	440.90	1776	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-3	-	7.00	103.13	468.26	1896	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-17	-	7.00	103.02	474.85	1925	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-5	-	7.00	102.92	480.98	1952	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-263-023-3	+	7.00	101.41	583.77	2409	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-894-997-49	-	7.00	97.67	943.33	4057	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-781-802-5	-	7.00	97.61	950.39	4090	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-694-078-5	-	7.00	97.61	950.39	4090	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-485-355B-45	+	7.00	95.66	1.2e+03	5368	
/cgn2_6/ptodata/2/ina/backfiles1.seq:5223424-5	-	7.00	95.64	1.2e+03	5379	Pate
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-477-451-9	+	7.00	95.35	1.3e+03	5599	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-477-451-13	-	7.00	95.35	1.3e+03	5599	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-470-260-4	+	7.00	94.95	1.3e+03	5925	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-471-491-4	+	7.00	94.95	1.3e+03	5925	
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-466-662-4	+	7.00	94.95	1.3e+03	5925	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-781-802-3	-	7.00	94.55	1.4e+03	6263	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-694-078-3	-	7.00	94.55	1.4e+03	6263	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-619-554-1	+	7.00	93.11	1.7e+03	7655	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-470-202-56	-	7.00	91.34	2.1e+03	9793	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-770-56	-	7.00	91.34	2.1e+03	9793	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-468-059-56	-	7.00	91.34	2.1e+03	9793	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-109-916-56	-	7.00	91.34	2.1e+03	9793	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-477-451-1	+	7.00	90.98	2.2e+03	10299	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-477-451-5	-	7.00	90.98	2.2e+03	10299	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-477-451-25	+	7.00	86.24	4.1e+03	19932	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-422-869-1	-	7.00	79.76	9.3e+03	49136	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2	-	7.00	47.49	4.1e+05	44037	
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1	-	7.00	47.47	4.1e+05	44115	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-912-129A-66	-	6.00	118.57	64.61	19	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-004-139B-10	-	6.00	118.20	67.73	20	
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-811-492-10	-	6.00	118.20	67.73	20	

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-411-706-3

seq_documentation_block:
; Sequence 3, Application US/08411706
; Patent No. 5789193
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; APPLICANT:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: Increased production of secreted
; TITLE OF INVENTION: proteins by recombinant eukaryotic cel
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,706
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FI 92 4494
; FILING DATE: 06-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Thomas C.
; REGISTRATION NUMBER: 36,989
; REFERENCE/DOCKET NUMBER: 32530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; STRAIN: X 2180-1B
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..885
US-08-411-706-3

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-10-048-197-2 x US-08-411-706-3/rev ..

```
 Align seg 1/1  to reverse of: US-08-411-706-3  from: 1  to: 885

      12 AlaLeuLeuThrAlaLeuLeuVal 19
         ||||||||||||||||||||||||
     779 GCACTCTTAACGGCCTTGTTGGTG 756

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-318-978-15

seq_documentation_block:
; Sequence 15, Application US/09318978A
; Patent No. 6245526
; GENERAL INFORMATION:
;  APPLICANT: Yue, Henry
;  APPLICANT: Kaser, Matthew, R.
;  APPLICANT: Baughn, Mariah, R.
;  TITLE OF INVENTION: LIPID METABOLISM TRANSCRIPTION FACTOR
;  FILE REFERENCE: PC-0004 US
;  CURRENT APPLICATION NUMBER: US/09/318,978A
;  CURRENT FILING DATE: 1999-05-26
;  NUMBER OF SEQ ID NOS: 27
;  SOFTWARE: PERL Program
; SEQ ID NO 15
;   LENGTH: 250
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 45
;   OTHER INFORMATION: a or g or c or t, unknown, or other
;   FEATURE:
;   NAME/KEY:
;   OTHER INFORMATION: 701252210H1
;  PUBLICATION INFORMATION:
US-09-318-978-15


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-09-318-978-15   ..

 Align seg 1/1  to: US-09-318-978-15  from: 1  to: 250

      15 ThrAlaLeuLeuValThrGly 21
         ||||||||||||||||||||
      85 ACAGCACTGTTGGTAACCGGT 105

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-747-221B-1

seq_documentation_block:
; Sequence 1, Application US/08747221B
; Patent No. 6063610
;  GENERAL INFORMATION:
;    APPLICANT:  Silver, Gary W.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  No. 6063610el Carboxylesterase Nucleic Acid
;    TITLE OF INVENTION:  Molecules, Proteins and Uses Thereof
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carol Talkington Verser, Ph.D.
;      ADDRESSEE:  Heska Corporation
;      STREET:  1825 Sharp Point Drive
;      CITY:  Fort Collins
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80525
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows, Version 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,221B
;      FILING DATE:  No. 6063610ember 12, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verser, Carol Talkington
;      REGISTRATION NUMBER:  37,459
;      REFERENCE/DOCKET NUMBER:  FC-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  970/493-7272
;      TELEFAX:  970/484-9505
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  401 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  92..400
;    FEATURE:
;      NAME/KEY:  Xaa = Ile, Thr Lys or Arg
;      LOCATION:  218
;    FEATURE:
;      NAME/KEY:  Xaa = Lys, Glu or Gln
;      LOCATION:  275, 329
;    FEATURE:
;      NAME/KEY:  Xaa = Asn, Tyr or Asp
;      LOCATION:  332
US-08-747-221B-1


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-747-221B-1/rev   ..

 Align seg 1/1  to reverse of: US-08-747-221B-1  from: 1  to: 401

      86 LeuSerLeuIleIleSerPhe 92
         ||||||||||||||||||||||
     156 TTATCATTAATCATTTCCTTC 136

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-747-221B-3

seq_documentation_block:
; Sequence 3, Application US/08747221B
; Patent No. 6063610
;  GENERAL INFORMATION:
;    APPLICANT:  Silver, Gary W.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  No. 6063610el Carboxylesterase Nucleic Acid
;    TITLE OF INVENTION:  Molecules, Proteins and Uses Thereof
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carol Talkington Verser, Ph.D.
;      ADDRESSEE:  Heska Corporation
;      STREET:  1825 Sharp Point Drive
;      CITY:  Fort Collins
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80525
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows, Version 7.0
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,221B
;      FILING DATE:  No. 6063610ember 12, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verser, Carol Talkington
;      REGISTRATION NUMBER:  37,459
;      REFERENCE/DOCKET NUMBER:  FC-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  970/493-7272
;      TELEFAX:  970/484-9505
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  401 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-747-221B-3


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-747-221B-3    ..

 Align seg 1/1  to: US-08-747-221B-3  from: 1  to: 401

      86 LeuSerLeuIleIleSerPhe 92
         ||||||||||||||||||||
     246 TTATCATTAATCATTTCCTTC 266

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-005-051-1

seq_documentation_block:
; Sequence 1, Application US/09005051
; Patent No. 6291222
;  GENERAL INFORMATION:
;    APPLICANT:  Silver, Gary W.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  No. 6291222el Carboxylesterase Nucleic Acid
;    TITLE OF INVENTION:  Molecules, Proteins and Uses Thereof
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carol Talkington Verser, Ph.D.
;      ADDRESSEE:  Heska Corporation
;      STREET:  1825 Sharp Point Drive
;      CITY:  Fort Collins
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80525
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows, Version 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/005,051
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/747,221
;      FILING DATE:  No. 6291222ember 12, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verser, Carol Talkington
;      REGISTRATION NUMBER:  37,459
;      REFERENCE/DOCKET NUMBER:  FC-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  970/493-7272
;      TELEFAX:  970/484-9505
```

```
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  401 nucleotides
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  92..400
;    FEATURE:
;      NAME/KEY:  Xaa = Ile, Thr Lys or Arg
;      LOCATION:  218
;    FEATURE:
;      NAME/KEY:  Xaa = Lys, Glu or Gln
;      LOCATION:  275, 329
;    FEATURE:
;      NAME/KEY:  Xaa = Asn, Tyr or Asp
;      LOCATION:  332
US-09-005-051-1


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-09-005-051-1/rev    ..

 Align seg 1/1  to reverse of: US-09-005-051-1  from: 1  to: 401

      86 LeuSerLeuIleIleSerPhe 92
         ||||||||||||||||||||
     156 TTATCATTAATCATTTCCTTC 136

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-005-051-3

seq_documentation_block:
; Sequence 3, Application US/09005051
; Patent No. 6291222
;  GENERAL INFORMATION:
;    APPLICANT:  Silver, Gary W.
;    APPLICANT:  Wisnewski, Nancy
;    TITLE OF INVENTION:  No. 6291222el Carboxylesterase Nucleic Acid
;    TITLE OF INVENTION:  Molecules, Proteins and Uses Thereof
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Carol Talkington Verser, Ph.D.
;      ADDRESSEE:  Heska Corporation
;      STREET:  1825 Sharp Point Drive
;      CITY:  Fort Collins
;      STATE:  Colorado
;      COUNTRY:  USA
;      ZIP:  80525
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  Windows 95
;      SOFTWARE:  WordPerfect for Windows, Version 7.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/005,051
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/747,221
;      FILING DATE:  No. 6291222ember 12, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Verser, Carol Talkington
;      REGISTRATION NUMBER:  37,459
;      REFERENCE/DOCKET NUMBER:  FC-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  970/493-7272
```

```
;       TELEFAX:  970/484-9505
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  401 nucleotides
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-09-005-051-3


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-09-005-051-3    ..

 Align seg 1/1  to: US-09-005-051-3  from: 1  to: 401

      86 LeuSerLeuIleIleSerPhe 92
         ||||||||||||||||||||
     246 TTATCATTAATCATTTCCTTC 266

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-367-198A-4

seq_documentation_block:
; Sequence 4, Application US/08367198A
; Patent No. 5627049
;  GENERAL INFORMATION:
;    APPLICANT:  Menart, Sandrine
;    APPLICANT:  Bolotin, Monique
;    TITLE OF INVENTION:  K. lactis RP28 Ribosomal Protein Gene
;    TITLE OF INVENTION:  Promoter and Use Thereof
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rhone-Poulenc Rorer Inc.
;      STREET:  500 Arcola Road, 3C43
;      CITY:  Collegeville
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19426-0107
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/367,198A
;      FILING DATE:  01-FEB-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR93/00695
;      FILING DATE:  06-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/08429
;      FILING DATE:  08-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, Julie K.
;      REGISTRATION NUMBER:  38,619
;      REFERENCE/DOCKET NUMBER:  ST92045-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610)454-3839
;      TELEFAX:  (610)454-3808
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  549 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
US-08-367-198A-4


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-367-198A-4/rev    ..

 Align seg 1/1  to reverse of: US-08-367-198A-4  from: 1  to: 549

       4 HisHisIleArgLeuThrIle 10
         ||||||||||||||||||||
     241 CATCACATCAGACTGACGATC 221

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-932-978-1

seq_documentation_block:
; Sequence 1, Application US/08932978
; Patent No. 5885804
;  GENERAL INFORMATION:
;    APPLICANT:  Zalacain, Magdalena
;    APPLICANT:  Brown, James R.
;    TITLE OF INVENTION:  NOVEL phoH
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dechert Price & Rhoads
;      STREET:  4000 Bell Atlantic Tower, 1717 Arch Stre
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  US
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/932,978
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dickinson, Todd Q
;      REGISTRATION NUMBER:  28,354
;      REFERENCE/DOCKET NUMBER:  GM0100
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-994-2252
;      TELEFAX:  215-994-2222
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  969 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-932-978-1


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-932-978-1/rev    ..

 Align seg 1/1  to reverse of: US-08-932-978-1  from: 1  to: 969
```

```
     86 LeuSerLeuIleIleSerPhe 92
        ||||||||||||||||||||
    317 TTATCTTTGATAATTTCTTTT 297

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-367-198A-1

seq_documentation_block:
; Sequence 1, Application US/08367198A
; Patent No. 5627049
;  GENERAL INFORMATION:
;    APPLICANT:  Menart, Sandrine
;    APPLICANT:  Bolotin, Monique
;    TITLE OF INVENTION:  K. lactis RP28 Ribosomal Protein Gene
;    TITLE OF INVENTION:  Promoter and Use Thereof
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rhone-Poulenc Rorer Inc.
;      STREET:  500 Arcola Road, 3C43
;      CITY:  Collegeville
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19426-0107
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/367,198A
;      FILING DATE:  01-FEB-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/FR93/00695
;      FILING DATE:  06-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/08429
;      FILING DATE:  08-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, Julie K.
;      REGISTRATION NUMBER:  38,619
;      REFERENCE/DOCKET NUMBER:  ST92045-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (610)454-3839
;      TELEFAX:  (610)454-3808
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  970 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Kluyveromyces lactis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  883..969
US-08-367-198A-1


alignment_scores:
           Quality:    7.00            Length:      7
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-367-198A-1/rev   ..

 Align seg 1/1  to reverse of: US-08-367-198A-1  from: 1  to: 970

       4 HisHisIleArgLeuThrIle 10
         ||||||||||||||||||||
     577 CATCACATCAGACTGACGATC 557

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-961-083-7

seq_documentation_block:
; Sequence 7, Application US/08961083
; Patent No. 6159469
;  GENERAL INFORMATION:
;    APPLICANT:  Choi et. al.
;    TITLE OF INVENTION:  Streptococcus pneumoniae Antigens and Vaccines
;    NUMBER OF SEQUENCES:  452
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,083
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  985 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-083-7


alignment_scores:
           Quality:    7.00            Length:      7
             Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-961-083-7/rev   ..

 Align seg 1/1  to reverse of: US-08-961-083-7  from: 1  to: 985

      42 IleIleLysGlyLysThrAsn 48
         ||||||||||||||||||||
     268 ATTATTAAGGGCAAAACCAAC 248

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-684-862-9

seq_documentation_block:
; Sequence 9, Application US/08684862
; Patent No. 5759541
;  GENERAL INFORMATION:
;    APPLICANT:  Bach, Alfred
;    APPLICANT:  Hillen, Heinz
;    APPLICANT:  Bialojan, Siegfried
;    TITLE OF INVENTION:  No. 5759541el Proteins, the Preparation and Use
;    TITLE OF INVENTION:  Thereof
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Keil & Weinkauf
```

```
;       STREET:  1101 Connecticut Avenue
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 5.25 inch, 360 Kb storage
;       COMPUTER:  IBM AT-compatible, 80286 processor
;       OPERATING SYSTEM:  MS-DOS version 5.0
;       SOFTWARE:  WordPerfect version 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/684,862
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/361,705
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  07/966,040
;       FILING DATE:  30-DEC-1992
;       APPLICATION NUMBER:  PCT/EP91/01361
;       FILING DATE:  19-JUL-1991
;   INFORMATION FOR SEQ ID NO:  9:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1333 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Agkistrodon rhodostoma
;     FEATURE:
;       LOCATION:  231 to 935
;       OTHER INFORMATION:  the coding region shown in (2)(ix)(B)
;       OTHER INFORMATION:  codes for the protein of SEQ ID NO: 4
US-08-684-862-9


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-684-862-9/rev    ..

 Align seg 1/1  to reverse of: US-08-684-862-9  from: 1  to: 1333

      87 SerLeuIleIleSerPheLeu 93
         ||||||||||||||||||||||
    1298 TCTCTTATTATTTCCTTTCTC 1278

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-13

seq_documentation_block:
; Sequence 13, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 1699
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria  E004
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (92)..(1594)
US-09-058-260-13


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-09-058-260-13/rev    ..

 Align seg 1/1  to reverse of: US-09-058-260-13  from: 1  to: 1699

      51 GluIleSerSerArgPheGly 57
         |||||||||||||||||||||
    1413 GAAATTAGCTCCCGCTTCGGA 1393

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-21

seq_documentation_block:
; Sequence 21, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
;   LENGTH: 1753
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E013
```

```
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (128)..(1630)
US-09-058-260-21


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-09-058-260-21/rev    ..

 Align seg 1/1  to reverse of: US-09-058-260-21  from: 1  to: 1753

      51 GluIleSerSerArgPheGly 57
         ||||||||||||||||||||
    1449 GAAATTAGCTCCCGCTTCGGA 1429

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-31

seq_documentation_block:
; Sequence 31, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 1756
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E027
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (130)..(1632)
US-09-058-260-31


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-09-058-260-31/rev    ..

 Align seg 1/1  to reverse of: US-09-058-260-31  from: 1  to: 1756

      51 GluIleSerSerArgPheGly 57
         ||||||||||||||||||||
    1451 GAAATTAGCTCCCGCTTCGGA 1431

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-23

seq_documentation_block:
; Sequence 23, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23
;   LENGTH: 1776
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E015
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (128)..(1630)
US-09-058-260-23


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:       0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-09-058-260-23/rev    ..

 Align seg 1/1  to reverse of: US-09-058-260-23  from: 1  to: 1776

      51 GluIleSerSerArgPheGly 57
         ||||||||||||||||||||
    1449 GAAATTAGCTCCCGCTTCGGA 1429

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-3

seq_documentation_block:
; Sequence 3, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
```

```
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1896
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E009
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (211)..(1713)
US-09-058-260-3


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-09-058-260-3/rev    ..

 Align seg 1/1  to reverse of: US-09-058-260-3  from: 1  to: 1896

      51 GluIleSerSerArgPheGly 57
         ||||||||||||||||||||
    1532 GAAATTAGCTCCCGCTTCGGA 1512

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-17

seq_documentation_block:
; Sequence 17, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 1925
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E008
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (127)..(1581)
US-09-058-260-17


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-09-058-260-17/rev    ..

 Align seg 1/1  to reverse of: US-09-058-260-17  from: 1  to: 1925

      51 GluIleSerSerArgPheGly 57
         ||||||||||||||||||||
    1400 GAAATTAGCTCCCGCTTCGGA 1380

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-5

seq_documentation_block:
; Sequence 5, Application US/09058260B
; Patent No. 6218167
; GENERAL INFORMATION:
;  APPLICANT: Allen, Larry
;  APPLICANT: Aikens, John
;  APPLICANT: Fonstein, Michael
;  APPLICANT: Vonstein, Veronika
;  APPLICANT: Demirjian, David
;  APPLICANT: Casadaban, Malcolm
;  TITLE OF INVENTION: Stable Biocatalysts for Ester Hydrolysis
;  FILE REFERENCE: 95-963-H
;  CURRENT APPLICATION NUMBER: US/09/058,260B
;  CURRENT FILING DATE: 1999-04-10
;  EARLIER APPLICATION NUMBER: 60/001,995
;  EARLIER FILING DATE: 1996-08-07
;  EARLIER APPLICATION NUMBER: 60/009,704
;  EARLIER FILING DATE: 1996-01-11
;  EARLIER APPLICATION NUMBER: 60/019,580
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: 08/694,078
;  EARLIER FILING DATE: 1996-08-08
;  EARLIER APPLICATION NUMBER: 08/781,802
;  EARLIER FILING DATE: 1997-01-10
;  EARLIER APPLICATION NUMBER: 08/827,810
;  EARLIER FILING DATE: 1997-04-11
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1952
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:cloned esterase
;   OTHER INFORMATION: gene from bacteria E011
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (197)..(1699)
US-09-058-260-5


alignment_scores:
```

```
           Quality:    7.00             Length:      7
             Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-09-058-260-5/rev   ..

 Align seg 1/1  to reverse of: US-09-058-260-5  from: 1  to: 1952

      51 GluIleSerSerArgPheGly 57
         |||||||||||||||||||||
    1518 GAAATTAGCTCCCGCTTCGGA 1498

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-263-023-3

seq_documentation_block:
; Sequence 3, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
;  APPLICANT: Uchimura, Kenji
;  APPLICANT: Muramatsu, Hideki
;  APPLICANT: Kadomatsu, Kenji
;  APPLICANT: Kannagi, Reiji
;  APPLICANT: Habuchi, Osami
;  APPLICANT: Muramatsu, Takashi
;  TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
;  TITLE OF INVENTION: DNA ENCODING THE SAME
;  FILE REFERENCE: TOYAM41.001AUS
;  CURRENT APPLICATION NUMBER: US/09/263,023
;  CURRENT FILING DATE: 1999-03-05
;  EARLIER APPLICATION NUMBER: JP 10-54007
;  EARLIER FILING DATE: 1998-03-05
;  EARLIER APPLICATION NUMBER: JP 10-177844
;  EARLIER FILING DATE: 1998-06-24
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 2409
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (390)...(1841)
US-09-263-023-3


alignment_scores:
           Quality:    7.00             Length:      7
             Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-09-263-023-3   ..

 Align seg 1/1  to: US-09-263-023-3  from: 1  to: 2409

      11 SerAlaLeuLeuThrAlaLeu 17
         |||||||||||||||||||||
     614 TCCGCACTCCTTACCGCCCTC 634

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-894-997-49

seq_documentation_block:
; Sequence 49, Application US/08894997A
; Patent No. 6270990
; GENERAL INFORMATION:
;  APPLICANT: Anderson, David J
;  APPLICANT: Schoenherr, Christopher J
;  TITLE OF INVENTION: NEURON-RESTRICTIVE SILENCER FACTOR
;  FILE REFERENCE: 17810-502 NRSF
;  CURRENT APPLICATION NUMBER: US/08/894,997A
;  CURRENT FILING DATE: 1998-01-06
;  EARLIER APPLICATION NUMBER: PCT/US96/02817
;  EARLIER FILING DATE: 1996-03-01
;  EARLIER APPLICATION NUMBER: 08/398,590
;  EARLIER FILING DATE: 1995-03-03
;  NUMBER OF SEQ ID NOS: 55
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
;   LENGTH: 4057
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: gene
;   LOCATION: (1)..(4057)
;   OTHER INFORMATION: Human NSRF
US-08-894-997-49


alignment_scores:
           Quality:    7.00             Length:      7
             Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-894-997-49/rev   ..

 Align seg 1/1  to reverse of: US-08-894-997-49  from: 1  to: 4057

      44 LysGlyLysThrAsnLysGln 50
         ||||||||||||||||||||
   4011 AAAGGAAAAACAAACAAACAA 3991

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-781-802-5

seq_documentation_block:
; Sequence 5, Application US/08781802
; Patent No. 5969121
;  GENERAL INFORMATION:
;    APPLICANT:  ALLEN, Larry
;    APPLICANT:  AIKENS, John
;    APPLICANT:  FONSTEIN, Michael
;    APPLICANT:  VONSTEIN, Veronika
;    APPLICANT:  DEMIRJIAN, David
;    APPLICANT:  CASADABAN, Malcolm
;    TITLE OF INVENTION:  Stable Biocatalysts for Ester Hydrolysis
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 S. Wacker Drive 32nd Floor
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,802
;      FILING DATE:  10-JAN-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/694,078
;      FILING DATE:  07-AUG-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/019,580
;      FILING DATE:  12-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/009,704
;      FILING DATE:  11-JAN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/001,995
;      FILING DATE:  01-AUG-1995
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Chao, Mark
;      REGISTRATION NUMBER:  37,293
;      REFERENCE/DOCKET NUMBER:  95,963-E
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-913-0001
;      TELEFAX:  312-913-0002
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4090 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  197..1699
;      OTHER INFORMATION:  /note= "E011 sequence of longest
;      OTHER INFORMATION:  open reading frame; other possible start codons ATG/met4;
;      OTHER INFORMATION:  TTG/leu7; GTG/val8; GTG/val15; GTG/val36; ATG/met62"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  197..1699
US-08-781-802-5


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-781-802-5/rev   ..

 Align seg 1/1  to reverse of: US-08-781-802-5  from: 1  to: 4090

      51 GluIleSerSerArgPheGly 57
         |||||||||||||||||||||
    1518 GAAATTAGCTCCCGCTTCGGA 1498

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-694-078-5

seq_documentation_block:
; Sequence 5, Application US/08694078
; Patent No. 6218163
;  GENERAL INFORMATION:
;    APPLICANT:  ALLEN, Larry
;    APPLICANT:  AIKENS, John
;    APPLICANT:  FONSTEIN, Michael
;    APPLICANT:  VONSTEIN, Veronika
;    APPLICANT:  DEMIRJIAN, David
;    APPLICANT:  CASADABAN, Malcolm
;    TITLE OF INVENTION:  Stable Biocatalysts for Ester Hydrolysis
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff, Ltd.
;      STREET:  300 S. Wacker Drive 7th Floor
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/694,078
;      FILING DATE:  07-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/019,580
;      FILING DATE:  12-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/009,704
;      FILING DATE:  10-JAN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/001,995
;      FILING DATE:  07-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Chao, Mark
;      REGISTRATION NUMBER:  37,293
;      REFERENCE/DOCKET NUMBER:  95,963-C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-913-0001
;      TELEFAX:  312-913-0002
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4090 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  197..1699
;      OTHER INFORMATION:  /note= "E011 sequence of longest
;      OTHER INFORMATION:  open reading frame; other possible start codons ATG/met4;
;      OTHER INFORMATION:  TTG/leu7; GTG/val8; GTG/val15; GTG/val36; ATG/met62"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  197..1699
US-08-694-078-5


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-694-078-5/rev   ..

 Align seg 1/1  to reverse of: US-08-694-078-5  from: 1  to: 4090

      51 GluIleSerSerArgPheGly 57
         |||||||||||||||||||||
    1518 GAAATTAGCTCCCGCTTCGGA 1498

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-485-355B-45

seq_documentation_block:
; Sequence 45, Application US/08485355B
; Patent No. 6177075
;   GENERAL INFORMATION:
;        APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
;        TITLE OF INVENTION: Insect Viruses and Their Uses in
;                            Protecting Plants
;        NUMBER OF SEQUENCES: 57
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
;             STREET: Four Embarcadero Center, Suite 3400
;             CITY: San Francisco
;             STATE: California
;             COUNTRY: United States
;             ZIP: 94111-4187
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/08/485,355B
;             FILING DATE: 07-Jun-1995
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
```

```
;             APPLICATION NUMBER: US 08/440,522
;             FILING DATE: 12-MAY-1995
;             APPLICATION NUMBER: US 08/089,372
;             FILING DATE: 08-JUL-1993
;             APPLICATION NUMBER: AU PL4081/92
;             FILING DATE: 14-AUG-1992
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Trecartin, Richard F.
;             REGISTRATION NUMBER: 31,801
;             REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (415) 781-1989
;             TELEFAX: (415) 398-3249
;             TELEX: 910 277299
;   INFORMATION FOR SEQ ID NO: 45:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 5368 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: unknown
;             TOPOLOGY: unknown
;        MOLECULE TYPE: DNA
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  4944..5162
;        SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-08-485-355B-45


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-485-355B-45   ..

 Align seg 1/1  to: US-08-485-355B-45  from: 1  to: 5368

      30 MetLysGluGlnAsnGlnGln 36
         ||||||||||||||||||||
    5190 ATGAAGGAACAAAATCAGCAG 5210

seq_name: /cgn2_6/ptodata/2/ina/backfiles1.seq:5223424-5

seq_documentation_block:
;Patent No. 5223424
;    APPLICANT: COCHRAN, MARK;CHIANG, CHRISTINA H.;MACDONALD,
;RICHARD D.
;    TITLE OF INVENTION: ATTENUATED HERPESVIRUSES AND
;HERPESVIRUSES WHICH INCLUDE FOREIGN DNA ENCODING AN AMINO
;ACID SEQUENCE
;    NUMBER OF SEQUENCES: 16
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/225,032
;      FILING DATE: 27-JUL-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 78,519
;      FILING DATE: 27-JUL-1987
;      APPLICATION NUMBER: 933,107
;      FILING DATE: 20-NOV-1986
;      APPLICATION NUMBER: 902,887
;      FILING DATE: 02-SEP-1986
;      APPLICATION NUMBER: 887,140
;      FILING DATE: 17-JUL-1986
;      APPLICATION NUMBER: 823,102
;      FILING DATE: 27-JAN-1986
;      APPLICATION NUMBER: 773,430
;      FILING DATE: 06-SEP-1985
;SEQ ID NO:5:
;      LENGTH: 5379
5223424-5

alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x 5223424-5/rev   ..

 Align seg 1/1  to reverse of: 5223424-5  from: 1  to: 5379

       8 LeuThrIleSerAlaLeuLeu 14
         ||||||||||||||||||||
     733 TTAACAATAAGTGCGTTGTTA 713

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-477-451-9

seq_documentation_block:
; Sequence 9, Application US/08477451
; Patent No. 5928865
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,451
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0335.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2708
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5599 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-477-451-9


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-477-451-9   ..

 Align seg 1/1  to: US-08-477-451-9  from: 1  to: 5599

      85 IleLeuSerLeuIleIleSer 91
         ||||||||||||||||||||
    3333 ATTTTGAGCTTGATCATTTCG 3353

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-477-451-13

seq_documentation_block:
```

```
; Sequence 13, Application US/08477451
; Patent No. 5928865
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,451
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0335.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2708
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5599 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-477-451-13


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-477-451-13/rev   ..

 Align seg 1/1  to reverse of: US-08-477-451-13  from: 1  to: 5599

      85 IleLeuSerLeuIleIleSer 91
         |||||||||||||||||||||
    2267 ATTTTGAGCTTGATCATTTCG 2247

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-470-260-4

seq_documentation_block:
; Sequence 4, Application US/08470260
; Patent No. 6077706
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    APPLICANT:  Bugnoli, Massimo
;    APPLICANT:  Telford, John
;    APPLICANT:  Macchia, Giovanni
;    APPLICANT:  Rappuoli, Rino
;    TITLE OF INVENTION:  Helicobacter Pylori Proteins Useful
;    TITLE OF INVENTION:  for Vaccines and Diagnostics
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,260
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/256,848
;      FILING DATE:  21-OCT-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0316.001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (510) 601-2708
;      TELEFAX:  (510) 655-3542
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5925 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-470-260-4


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-470-260-4   ..

 Align seg 1/1  to: US-08-470-260-4  from: 1  to: 5925

      90 IleSerPheLeuTrpValLys 96
         |||||||||||||||||||||
    4801 ATCTCATTTTTGTGGGTAAAA 4821

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-471-491-4

seq_documentation_block:
; Sequence 4, Application US/08471491B
; Patent No. 6090611
; GENERAL INFORMATION:
;  APPLICANT: Covacci, Antonello
;  APPLICANT: Bugnoli, Massimo
;  APPLICANT: Telford, John
;  APPLICANT: Macchia, Giovanni
;  APPLICANT: Rappuoli, Rino
;  TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
;  TITLE OF INVENTION: Diagnostics
;  FILE REFERENCE: CHIR0044
;  CURRENT APPLICATION NUMBER: US/08/471,491B
;  CURRENT FILING DATE: 1995-06-06
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 5925
;   TYPE: DNA
;   ORGANISM: Helicobacter pylori
US-08-471-491-4


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000
```

```
alignment_block:
 US-10-048-197-2 x US-08-471-491-4   ..

 Align seg 1/1  to: US-08-471-491-4  from: 1  to: 5925

      90 IleSerPheLeuTrpValLys 96
         ||||||||||||||||||||
    4801 ATCTCATTTTTGTGGGTAAAA 4821

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-466-662-4

seq_documentation_block:
; Sequence 4, Application US/08466662B
; Patent No. 6130059
; GENERAL INFORMATION:
;  APPLICANT: Covacci, Antonello
;  APPLICANT: Bugnoli, Massimo
;  APPLICANT: Telford, John
;  APPLICANT: Macchia, Giovanni
;  APPLICANT: Rappuoli, Rino
;  TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
;  TITLE OF INVENTION: Diagnostics
;  FILE REFERENCE: CHIR0057
;  CURRENT APPLICATION NUMBER: US/08/466,662B
;  CURRENT FILING DATE: 1995-06-06
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 5925
;   TYPE: DNA
;   ORGANISM: Helicobacter pylori
US-08-466-662-4


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-466-662-4   ..

 Align seg 1/1  to: US-08-466-662-4  from: 1  to: 5925

      90 IleSerPheLeuTrpValLys 96
         ||||||||||||||||||||
    4801 ATCTCATTTTTGTGGGTAAAA 4821

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-781-802-3

seq_documentation_block:
; Sequence 3, Application US/08781802
; Patent No. 5969121
;  GENERAL INFORMATION:
;    APPLICANT:  ALLEN, Larry
;    APPLICANT:  AIKENS, John
;    APPLICANT:  FONSTEIN, Michael
;    APPLICANT:  VONSTEIN, Veronika
;    APPLICANT:  DEMIRJIAN, David
;    APPLICANT:  CASADABAN, Malcolm
;    TITLE OF INVENTION:  Stable Biocatalysts for Ester Hydrolysis
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 S. Wacker Drive 32nd Floor
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,802
;      FILING DATE:  10-JAN-1997
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/694,078
;      FILING DATE:  07-AUG-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/019,580
;      FILING DATE:  12-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/009,704
;      FILING DATE:  11-JAN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/001,995
;      FILING DATE:  01-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Chao, Mark
;      REGISTRATION NUMBER:  37,293
;      REFERENCE/DOCKET NUMBER:  95,963-E
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-913-0001
;      TELEFAX:  312-913-0002
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6263 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2314..3816
;      OTHER INFORMATION:  /note= "E009 sequence with longest
;      OTHER INFORMATION:  open reading frame; possible other start codons are ATG/me
;      OTHER INFORMATION:  TTG/leu7; GTG/val8; GTG/val15; GTG/val36"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  2314..3816
US-08-781-802-3


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-781-802-3/rev   ..

 Align seg 1/1  to reverse of: US-08-781-802-3  from: 1  to: 6263

      51 GluIleSerSerArgPheGly 57
         ||||||||||||||||||||
    3635 GAAATTAGCTCCCGCTTCGGA 3615

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-694-078-3

seq_documentation_block:
; Sequence 3, Application US/08694078
; Patent No. 6218163
;  GENERAL INFORMATION:
;    APPLICANT:  ALLEN, Larry
;    APPLICANT:  AIKENS, John
;    APPLICANT:  FONSTEIN, Michael
;    APPLICANT:  VONSTEIN, Veronika
;    APPLICANT:  DEMIRJIAN, David
;    APPLICANT:  CASADABAN, Malcolm
;    TITLE OF INVENTION:  Stable Biocatalysts for Ester Hydrolysis
;    NUMBER OF SEQUENCES:  8
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff, Ltd.
;      STREET:  300 S. Wacker Drive 7th Floor
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/694,078
;      FILING DATE:  07-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/019,580
;      FILING DATE:  12-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/009,704
;      FILING DATE:  10-JAN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/001,995
;      FILING DATE:  07-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Chao, Mark
;      REGISTRATION NUMBER:  37,293
;      REFERENCE/DOCKET NUMBER:  95,963-C
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-913-0001
;      TELEFAX:  312-913-0002
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6263 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2314..3816
;      OTHER INFORMATION:  /note= "E009 sequence with longest
;      OTHER INFORMATION:  open reading frame; possible other start codons are ATG/met4;
;      OTHER INFORMATION:  TTG/leu7; GTG/val8; GTG/val15; GTG/val36"
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  2314..3816
US-08-694-078-3


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-694-078-3/rev   ..

 Align seg 1/1  to reverse of: US-08-694-078-3  from: 1  to: 6263

      51 GluIleSerSerArgPheGly 57
         |||||||||||||||||||||
    3635 GAAATTAGCTCCCGCTTCGGA 3615

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-619-554-1

seq_documentation_block:
; Sequence 1, Application US/08619554
; Patent No. 5821353
;  GENERAL INFORMATION:
;    APPLICANT:  DOUGLAS, Cameron M.
;    APPLICANT:  CHREBET, Gary L.
;    APPLICANT:  CLEMAS, Joseph
;    APPLICANT:  EL-SHERBEINI, Mohammed
;    APPLICANT:  FOOR, Forrest
;    APPLICANT:  KAHN, Jennifer,
;    APPLICANT:  KELLY, Rosemarie, - PARENT, S.A.
;    APPLICANT:  MARRINAN, Jean, - RAMADAN, N.M.
;    APPLICANT:  MORIN, Nancy, - REGISTER, E.A
;    APPLICANT:  ONISHI, Janet, - SHEI, Gan-Ju
;    TITLE OF INVENTION:  DNA ENCODING 1,3 BETA-D GLUCAN
;    TITLE OF INVENTION:  SYNTHASE SUBUNITS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  JOSEPH A. COPPOLA - MERCK & CO., INC.
;      STREET:  126 EAST LINCOLN AVENUE - P.O. BOX 2000
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/619,554
;      FILING DATE:  01-AUG-1996
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  COPPOLA, JOSEPH A
;      REGISTRATION NUMBER:  38,413
;      REFERENCE/DOCKET NUMBER:  19104PI
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  732-594-6734
;      TELEFAX:  732-594-4720
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7655 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-619-554-1


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-619-554-1   ..

 Align seg 1/1  to: US-08-619-554-1  from: 1  to: 7655

      12 AlaLeuLeuThrAlaLeuLeu 18
         ||||||||||||||||||||
     496 GCCCTCCTTACTGCTCTCCTT 516

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-470-202-56

seq_documentation_block:
; Sequence 56, Application US/08470202
; Patent No. 5759808
;  GENERAL INFORMATION:
;    APPLICANT:  Guertler, Lutz G.
;    APPLICANT:  Eberle, Josef
;    APPLICANT:  Brunn, Albrecht v.
;    APPLICANT:  Knapp, Stefan
;    APPLICANT:  Hauser, Hans-Peter
```

```
;    TITLE OF INVENTION:  Retrovirus from the HIV Group and Its
;    TITLE OF INVENTION:  Use
;    NUMBER OF SEQUENCES:  63
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/470,202
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/132,653
;      FILING DATE:  05-OCT-1993
;      APPLICATION NUMBER:  DE P 42 33 646.5
;      FILING DATE:  06-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 42 35 718.7
;      FILING DATE:  22-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 42 44 541.8
;      FILING DATE:  30-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 43 18 186.4
;      FILING DATE:  01-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Michael J. Blake
;      REGISTRATION NUMBER:  37,096
;      REFERENCE/DOCKET NUMBER:  05495-0001-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-470-202-56


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-470-202-56/rev   ..

 Align seg 1/1  to reverse of: US-08-470-202-56  from: 1  to: 9793

      11 SerAlaLeuLeuThrAlaLeu 17
         ||||||||||||||||||||
    6097 TCAGCGTTACTTACTGCTCTG 6077

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-471-770-56

seq_documentation_block:
; Sequence 56, Application US/08471770
; Patent No. 5770427
;  GENERAL INFORMATION:
;    APPLICANT:  Guertler, Lutz G.
;    APPLICANT:  Eberle, Josef
;    APPLICANT:  Brunn, Albrecht v.
;    APPLICANT:  Knapp, Stefan
;    APPLICANT:  Hauser, Hans-Peter
;    TITLE OF INVENTION:  Retrovirus from the HIV Group and Its
;    TITLE OF INVENTION:  Use
;    NUMBER OF SEQUENCES:  63
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,770
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/132,653
;      FILING DATE:  05-OCT-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 42 33 646.5
;      FILING DATE:  06-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 42 35 718.7
;      FILING DATE:  22-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 42 44 541.8
;      FILING DATE:  30-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 43 18 186.4
;      FILING DATE:  01-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carol P. Einaudi
;      REGISTRATION NUMBER:  32,220
;      REFERENCE/DOCKET NUMBER:  05495-0001-03000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-471-770-56


alignment_scores:
            Quality:    7.00            Length:      7
              Ratio:  1.000              Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-471-770-56/rev   ..

 Align seg 1/1  to reverse of: US-08-471-770-56  from: 1  to: 9793

      11 SerAlaLeuLeuThrAlaLeu 17
         ||||||||||||||||||||
    6097 TCAGCGTTACTTACTGCTCTG 6077

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-468-059-56

seq_documentation_block:
```

```
; Sequence 56, Application US/08468059
; Patent No. 5840480
;  GENERAL INFORMATION:
;    APPLICANT:  Guertler, Lutz G.
;    APPLICANT:  Eberle, Josef
;    APPLICANT:  Brunn, Albrecht v.
;    APPLICANT:  Knapp, Stefan
;    APPLICANT:  Hauser, Hans-Peter
;    TITLE OF INVENTION:  Retrovirus from the HIV Group and Its
;    TITLE OF INVENTION:  Use
;    NUMBER OF SEQUENCES:  63
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow,. Garrett &
;      ADDRESSEE:  Dunner
;      STREET:  1300 I Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/468,059
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/132,653
;      FILING DATE:  05-OCT-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 42 33 646.5
;      FILING DATE:  06-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 42 35 718.7
;      FILING DATE:  22-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 42 44 541.8
;      FILING DATE:  30-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P 43 18 186.4
;      FILING DATE:  01-JUN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Carol P. Einaudi
;      REGISTRATION NUMBER:  32,220
;      REFERENCE/DOCKET NUMBER:  05495-0001-02000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-408-4000
;      TELEFAX:  202-408-4400
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9793 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-468-059-56


alignment_scores:
            Quality:    7.00              Length:      7
              Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-468-059-56/rev    ..

 Align seg 1/1  to reverse of: US-08-468-059-56  from: 1  to: 9793

      11 SerAlaLeuLeuThrAlaLeu 17
         ||||||||||||||||||||
```

```
    6097 TCAGCGTTACTTACTGCTCTG 6077

 seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-109-916-56

 seq_documentation_block:
 ; Sequence 56, Application US/09109916
 ; Patent No. 6277561
 ; GENERAL INFORMATION:
 ;  APPLICANT: Guertler, Lutz G.
 ;  APPLICANT: Eberle, Josef
 ;  APPLICANT: Brunn, Albrecht V.
 ;  APPLICANT: Knapp, Stefan
 ;  APPLICANT: Hauser, Hans-Peter
 ;  TITLE OF INVENTION: RETROVIRUS FROM HIV GROUP AND ITS USE
 ;  FILE REFERENCE: 05495.0001-04
 ;  CURRENT APPLICATION NUMBER: US/09/109,916
 ;  CURRENT FILING DATE: 1998-07-02
 ;  EARLIER APPLICATION NUMBER: DE P 42 33 646.5
 ;  EARLIER FILING DATE: 1992-10-06
 ;  EARLIER APPLICATION NUMBER: DE P 42 35 718.7
 ;  EARLIER FILING DATE: 1992-10-22
 ;  EARLIER APPLICATION NUMBER: DE P 42 44 541.8
 ;  EARLIER FILING DATE: 1992-12-30
 ;  EARLIER APPLICATION NUMBER: DE P 43 18 186.4
 ;  EARLIER FILING DATE: 1993-06-01
 ;  NUMBER OF SEQ ID NOS: 67
 ;  SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 56
 ;   LENGTH: 9793
 ;   TYPE: DNA
 ;   ORGANISM: Human immunodeficiency virus
 US-09-109-916-56


 alignment_scores:
             Quality:    7.00              Length:      7
               Ratio:  1.000                Gaps:      0
  Percent Similarity: 100.000   Percent Identity: 100.000

 alignment_block:
  US-10-048-197-2 x US-09-109-916-56/rev    ..

  Align seg 1/1  to reverse of: US-09-109-916-56  from: 1  to: 9793

       11 SerAlaLeuLeuThrAlaLeu 17
          ||||||||||||||||||||
     6097 TCAGCGTTACTTACTGCTCTG 6077

 seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-477-451-1

 seq_documentation_block:
 ; Sequence 1, Application US/08477451
 ; Patent No. 5928865
 ;  GENERAL INFORMATION:
 ;    APPLICANT:  Covacci, Antonello
 ;    TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
 ;    NUMBER OF SEQUENCES:  46
 ;    CORRESPONDENCE ADDRESS:
 ;      ADDRESSEE:  Chiron Corporation
 ;      STREET:  4560 Horton Street
 ;      CITY:  Emeryville
 ;      STATE:  CA
 ;      COUNTRY:  USA
 ;      ZIP:  94608-2916
 ;    COMPUTER READABLE FORM:
 ;      MEDIUM TYPE:  Floppy disk
 ;      COMPUTER:  IBM PC compatible
 ;      OPERATING SYSTEM:  PC-DOS/MS-DOS
 ;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
 ;    CURRENT APPLICATION DATA:
 ;      APPLICATION NUMBER:  US/08/477,451
 ;      FILING DATE:  07-JUN-1995
 ;      CLASSIFICATION:  435
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0335.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2708
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10299 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-477-451-1


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-477-451-1   ..

 Align seg 1/1  to: US-08-477-451-1  from: 1  to: 10299

      90 IleSerPheLeuTrpValLys 96
         ||||||||||||||||||||
    9175 ATCTCATTTTTGTGGGTAAAA 9195

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-477-451-5

seq_documentation_block:
; Sequence 5, Application US/08477451
; Patent No. 5928865
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,451
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0335.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2708
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10299 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-477-451-5


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-477-451-5/rev   ..

 Align seg 1/1  to reverse of: US-08-477-451-5  from: 1  to: 10299

      90 IleSerPheLeuTrpValLys 96
         ||||||||||||||||||||
    1125 ATCTCATTTTTGTGGGTAAAA 1105

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-477-451-25

seq_documentation_block:
; Sequence 25, Application US/08477451
; Patent No. 5928865
;  GENERAL INFORMATION:
;    APPLICANT:  Covacci, Antonello
;    TITLE OF INVENTION:  Helicobacter Pylori Cagi Region
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Chiron Corporation
;      STREET:  4560 Horton Street
;      CITY:  Emeryville
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94608-2916
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,451
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  McClung, Barbara G.
;      REGISTRATION NUMBER:  33,113
;      REFERENCE/DOCKET NUMBER:  0335.002
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-601-2708
;      TELEFAX:  510-655-3542
;  INFORMATION FOR SEQ ID NO:  25:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19932 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-477-451-25


alignment_scores:
            Quality:    7.00             Length:      7
              Ratio:  1.000               Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-08-477-451-25   ..

 Align seg 1/1  to: US-08-477-451-25  from: 1  to: 19932

      85 IleLeuSerLeuIleIleSer 91
         ||||||||||||||||||||
    6890 ATTTTGAGCTTGATCATTTCG 6910

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-422-869-1

seq_documentation_block:
```

```
; Sequence 1, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
;  APPLICANT: POLONSKY, KENNETH S.
;  APPLICANT: HORIKAWA, YUKIO
;  APPLICANT: ODA, NAOHISA
;  APPLICANT: COX, NANCY J.
;  APPLICANT: SREENAN, SEAMUS
;  APPLICANT: ZHOU, YUN-PING
;  APPLICANT: OTANI, KENICHI
;  APPLICANT: HANIS, CRAIG L.
;  APPLICANT: BELL, GRAEME I.
;  TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
;  FILE REFERENCE: ARCD:307
;  CURRENT APPLICATION NUMBER: US/09/422,869
;  CURRENT FILING DATE: 1999-10-21
;  EARLIER APPLICATION NUMBER: 60/134,175
;  EARLIER FILING DATE: 1999-05-13
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 49136
;   TYPE: DNA
;   ORGANISM: Human
US-09-422-869-1


alignment_scores:
           Quality:    7.00              Length:      7
             Ratio:  1.000                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-10-048-197-2 x US-09-422-869-1/rev   ..

 Align seg 1/1  to reverse of: US-09-422-869-1  from: 1  to: 49136

       8 LeuThrIleSerAlaLeuLeu 14
         ||||||||||||||||||||||
   46568 CTAACCATATCAGCCCTCCTG 46548
```